

89966

STIC-Biotech/ChemLib

Fr m: Chan, Christina
Sent: Wednesday, March 26, 2003 11:08 AM
To: Ford, Vanessa; STIC-Biotech/ChemLib
Subject: RE: In Re:09884465

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
308-3973
CM-1, 9B19

-----Original Message-----

From: Ford, Vanessa
Sent: Wednesday, March 26, 2003 10:57 AM
To: Chan, Christina
Subject: In Re:09884465

Please search SEQ ID NO:332. Please include interference searches. Please rush.

Vanessa L. Ford
Biotechnology Patent Examiner
Office: CM1 8A16
Mailbox: CM1 8E12
Phone: 703.308.4735

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 3/27
Date Completed: 3/28
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:53:17 ; Search time 49 Seconds
(without alignments)
1777.506 Million cell updates/sec

Title: US-09-884-465a-332
Perfect score: 4728
Sequence: 1 MQIYTDDEIQVAKLAGRYT.....IELRLPSGEVKKNLSDIFA 906
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3357	71.0	1039	2	H95115	conserved hypotet
2	3356	71.0	1039	2	D97985	hypothetical prote
3	1307.5	27.7	839	2	G95115	conserved hypotet
4	1245	26.3	853	2	C97985	hypothetical prote
5	946.5	20.0	819	2	B95136	conserved domain p
6	946	20.0	855	2	D98004	histidine Motif-Co
7	636	13.5	802	2	C95136	conserved domain p
8	636	13.5	828	2	E98004	hypothetical prote
9	223.5	4.7	822	2	T46758	hypothetical 92.4k
10	202.5	4.3	2269	2	T86777	rhodry protein -
11	201.5	4.3	1873	2	T30944	surface protein pr
12	201	4.3	1790	2	S67593	transport protein
13	201	4.3	2004	2	F95133	immunoglobulin A1
14	197	4.2	1478	2	S20117	protein kinase BCK
15	197	4.2	3488	2	T34418	hypothetical prote
16	194.5	4.1	1134	2	A60234	Iga Fc receptor pr
17	194.5	4.1	1164	1	FC50AG	Iga Fc receptor pr
18	193.5	4.1	1125	2	E90598	membrane nuclease,
19	192	4.1	1871	2	D96796	probable heat shoc
20	190	4.0	4688	2	F92885	hypothetical prote
21	189.5	4.0	2748	2	S7976	nuclear migration
22	189.5	4.0	5005	2	F92884	hypothetical prote
23	189.5	4.0	5327	2	T13564	microtubule-associ
24	186	3.9	1640	2	A24594	probable major sur
25	186	3.9	1658	2	S55101	hypothetical prote
26	185.5	3.9	1983	2	G96643	hypothetical prote
27	185.5	3.9	2109	2	E99066	protein H05009.1
28	185.5	3.9	2109	2	T33247	hypothetical prote
29	185	3.9	2485	1	H71621	serine/threonine-s

BEST AVAILABLE COPY

RESULT 1

H95115
conserved hypothetical protein SP1004 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95115
R:Retcelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequences of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; M01D:21357209; PMID:11463916
A:Accession: H95115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:g14972477; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1004

Query Match 71.0%; Score 3357; DB 2; Length 1039;

Best Local Similarity 72.3%; Pred. No. 2,1e-157; Indels 128; Gaps 12;

Matches 687; Conservative 40; Mismatches 95;

QY	7	DDEIQVAKLAGRYTTEDGYIF	-----DTS-----	WIKKDSLSERAAAAQA	47
DB	168	NSNVAVARSGRYTTNDGVFNPAIDTGNAYIVPHGHVHYIPKSDLSASELAAKA	227		
QY	48	YAKERGLTP-----PSTDHQDSNGTEAKCA-----	-----EATYNRVKAARKVPL	87	
DB	228	HLAGKNQPSQLSYSTASNDNTQSVAKGTSKPANKSENLAQLKLYDSPSAQRYSES	287		
QY	88	DRMPYMLQTVTKNGSLIIPSYDVHYHNKFEWFDGLYEAPKGYSLDGLATVKYVPEP	147		
DB	288	DGLVFDPAKIIISPTNGVAIPHGDHYHFIYS-----KLSALEEKIANN	---VPI	334	
QY	148	RNASDHVRKNKADQSKPDEKDEHDEVSEPTPESEKENHAGLNPSADNLYKPSDTDEE	207		
DB	335	SGTGSTVSN-----AKPNEVV-----SSLSGLSSN	---PSSLTTS	367	
QY	208	TEEEAEDTTDEAIEPCTPSIRONAMETLGLKSSLLGKDNNTTSAEVDLSLALKEQ	267		
DB	368	KE-----LSSASDGYIFNPKD-----	---IVEETA	389	
QY	268	RAPI-----OGPOIGOPTLPNNSLATPSPSLPNPCTSHKHEEDGYGFDANFI	316		
DB	390	TAIYVHGDFHYIPKSNQIGOPTLPNNSLATPSPSLPNPCTSHKHEEDGYGFDANFI	449		
QY	317	IAEDSGFGVMSHGDSNHYFFKKDLTEEQKAAQKHLKEVKTSHNGLDLSLSEHQDYPGNA	376		

Iga-specific metal
hypothetical prote
protein-tyrosine-p
cytadherence-acces
repeat organellar
mtprd protein - mo
FMTB protein [impo
collagen adhesin -
glutamate rich pro
major merozolite su
hypothetical prote
hypothetical prote
hypothetical prote
zinc metalloprotel
150K mating aggreg
SEC16 protein - ye

ALIGNMENTS

30	184.5	3.9	1963	2	B98002
31	183	3.9	2218	2	B84683
32	182	3.8	2450	2	S71625
33	181.5	3.8	1139	1	E64234
34	181	3.8	1939	2	T18372
35	181	3.8	1979	2	JW0059
36	181	3.8	2481	2	D90011
37	180.5	3.8	1185	2	A42404
38	180	3.8	1271	2	A45555
39	180	3.8	1639	2	S05603
40	179.5	3.8	2447	2	T18670
41	179	3.8	1969	2	T38495
42	178.5	3.8	763	2	T08929
43	177.5	3.8	1881	2	H95076
44	177	3.7	1305	2	H41662
45	176.5	3.7	2195	2	S61103

Db 450 IADESGFVMSHGDNHFFKDLTEQIKAAQKHLSEVKTSHNGDLSLSSHEQDYPNSA 509
 QY 377 KEMKLDKKIEEKIAGIKOYGVKRESIVNNEKNAIIPSGDHHADPIDEHKPVGIGH 436
 Db 510 KEMKLDKKIEEKIAGIKOYGVKRESIVNNEKNAIIPSGDHHADPIDEHKPVGIGH 569
 QY 437 SHSNYELFPKPEGVAKKGNKYITGEELTNVNNLKNSTFNNQFTLANGOKRVSFSP 496
 Db 570 SHSNYELFPKPEGVAKKGNKYITGEELTNVNNLKNSTFNNQFTLANGOKRVSFSP 629
 QY 497 ELEKLGINMLVKLITPDGKLVKSGVFGVGVGNIANFELDPYLPQGTFTYIASKD 556
 Db 630 ELEKLGINMLVKLITPDGKLVKSGVFGVGVGNIANFELDPYLPQGTFTYIASKD 689
 QY 557 YPEVSDGTFTVPTSLAYKMASQITFYFPHAGDTYLRVNPQFAPVPGTDLVRVDFEFG 616
 Db 690 YPEVSDGTFTVPTSLAYKMASQITFYFPHAGDTYLRVNPQFAPVPGTDLVRVDFEFG 749
 QY 617 NAYLENNYKVGEEKIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEPPILEK 676
 Db 750 NAYLENNYKVGEEKIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEPPILEK 809
 QY 677 TDKPSILPQFKNRAQENKLDKVEEPTSEKVEKELSETGNSTNLEEYPTVDPV 736
 Db 810 TDKPSILPQFKNRAQENKLDKVEEPTSEKVEKELSETGNSTNLEEYPTVDPV 869
 QY 737 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPS 796
 Db 870 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPS 929
 QY 797 KYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGVNGSDPMLDPALE 856
 Db 930 KYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGVNGSDPMLDPALE 989
 QY 857 AVDPVOEKLEKFTASYGLDLSVFNMDGTIELRLPSGEVYKKNLSDFIA 906
 Db 990 AVDPVOEKLEKFTASYGLDLSVFNMDGTIELRLPSGEVYKKNLSDLIA 1039

RESULT 2
 D97985
 hypothetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C:Accession: D97985
 R: Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
 r, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A: Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A: Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A: Reference number: A97872; MUID: 21429245; PMID: 11544234
 A: Accession: D97985
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-1039 <KUR>
 A: Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:gl5458515; GSPDB:GN00174
 A: Genes: phtE

Query Match 71.0%; Score 3356; DB 2; Length 1039;
 Best Local Similarity 72.0%; Pred. No. 2.4e-157;
 Matches 684; Conservative 43; Mismatches 95; Indels 128; Gaps 11;
 QY 7 DDEIQVAKLAGKTTEDGYTF-----DTS-----WIKKLSSEARRAAQA 47
 Db 168 NSNVAVARSQRTTNDGVFNADIEDTGNAYIVPHGHHYIPKPSLSASLAAKA 227
 QY 48 YAKEKGLTP-----PSTDHQDSGNTAKGA-----PAIYNKRAAKKVEL 87
 Db 228 HLAGKNQPSQLSYSTABDNNTQSVAKGTSKPNKSENQSLKELDPSAQRYSES 287
 QY 88 DRMPYNLIQYVEKNSLIIPSDYDHNIFKFWDEGLYAPKGYSLDILATVRYTYEP 147

Db 288 DGLVDFPAKITSIRTPNGVAIPHGHDYHFIPYS-----KLSALEKIA----- 329
 QY 148 RNASDRVKKNAQDQSKPDDEKHEDEVSEPTHPESDEKENHAGLNPSADNLYKPSDTSEE 207
 Db 330 RRVPISTGTVSTNAKPNEV-----SSLSLSN-----PSSLTYS 367
 QY 208 TEBAEDTTDEAIPGTPSIRQNAETLGLKSSLLGTRDNNNTISAEVDSLALAKESQ 267
 Db 368 KE-----LSSASDGYIFNPKD-----IVEETA 389
 QY 268 PAPI-----OGPOIGOPTLPNNSLATPSPINPGTSHKHEEDYGFDDANRI 316
 Db 390 TAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPINPGTSHKHEEDYGFDDANRI 449
 QY 317 IADESGFVMSHGDSNHYFFKDLTEQIKAAQKHLSEVKTSHNGDLSLSSHEQDYPNSA 376
 Db 450 IADESGFVMSHGDSNHYFFKDLTEQIKAAQKHLSEVKTSHNGDLSLSSHEQDYPNSA 509
 QY 377 KEMKLDKKIEEKIAGIKOYGVKRESIVNNEKNAIIPSGDHHADPIDEHKPVGIGH 436
 Db 510 KEMKLDKKIEEKIAGIKOYGVKRESIVNNEKNAIIPSGDHHADPIDEHKPVGIGH 569
 QY 437 SHSNYELFPKPEGVAKKGNKYITGEELTNVNNLKNSTFNNQFTLANGOKRVSFSP 496
 Db 570 SHSNYELFPKPEGVAKKGNKYITGEELTNVNNLKNSTFNNQFTLANGOKRVSFSP 629
 QY 497 ELEKLGINMLVKLITPDGKLVKSGVFGVGVGNIANFELDPYLPQGTFTYIASKD 556
 Db 630 ELEKLGINMLVKLITPDGKLVKSGVFGVGVGNIANFELDPYLPQGTFTYIASKD 689
 QY 557 YPEVSDGTFTVPTSLAYKMASQITFYFPHAGDTYLRVNPQFAPVPGTDLVRVDFEFG 616
 Db 690 YPEVSDGTFTVPTSLAYKMASQITFYFPHAGDTYLRVNPQFAPVPGTDLVRVDFEFG 749
 QY 617 NAYLENNYKVGEEKIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEPPILEK 676
 Db 750 NAYLENNYKVGEEKIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEPPILEK 809
 QY 677 TDKPSILPQFKNRAQENKLDKVEEPTSEKVEKELSETGNSTNLEEYPTVDPV 736
 Db 810 TDKPSILPQFKNRAQENKLDKVEEPTSEKVEKELSETGNSTNLEEYPTVDPV 869
 QY 737 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPS 796
 Db 870 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPS 929
 QY 797 KYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGVNGSDPMLDPALE 856
 Db 930 KYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGVNGSDPMLDPALE 989
 QY 857 AVDPVOEKLEKFTASYGLDLSVFNMDGTIELRLPSGEVYKKNLSDFIA 906
 Db 990 AVDPVOEKLEKFTASYGLDLSVFNMDGTIELRLPSGEVYKKNLSDLIA 1039

RESULT 3
 G95115
 conserved hypothetical protein SP1003 [imported] - Streptococcus pneumoniae (strain T
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: G95115
 R: Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
 son, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A: Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
 A: Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
 A: Reference number: A95000; MUID: 21357209; PMID: 11463916
 A: Accession: G95115
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-839 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75120.1; PID:gl4972476; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SP1003
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 27.7%; Score 1307.5; DB 2; Length 839;
 Best Local Similarity 81.7%; Pred. No. 9.1e-57;
 Matches 263; Conservative 2; Mismatches 6; Indels 51; Gaps 4;

QY 2 QITYTDEIQVAKLAGKYTTEDGYIFD-----TSWIKKDSLEAER 42
 DB 518 QITYTDEIQVAKLAGKYTTEDGYIFDPRDITSDGDAYVTPHMTSHWIKKDSLEAER 577

QY 43 AAAQAYAKEKGLTPPSTDHODSGNTEAKAEAIYNRVKAARKVPLDRMPYINQYTYVEKN 102
 DB 578 AAAQAYAKEKGLTPPSTDHODSGNTEAKAEAIYNRVKAARKVPLDRMPYINQYTYVEKN 637

QY 103 GSLIIPSYDHYHNKFEWDEGLYEAPKGYSLDILLATVYKYYE-----PR-----NAS 151
 DB 638 GSLIIPHYDHYHNKFEWDEGLYEAPKGYTLEDLLATVYKYYVEHNPHERPHSDNGFGNAS 697

QY 152 DHVRKNKADQSKPDEKDEHSEVTPHPESEKENHAGLNPSADNLYKPSYDTDETEEE 211
 DB 698 DHVRKNKVDQSKPDEKDEHSEVTPHPESEKENHAGLNPSADNLYKPSYDTDETEEE 757

QY 212 AEDTDEAEIPIGT-----PSIRQNAMETLGLKSSLLGTTKDN 250
 DB 758 AEDTDEAEIPIQVENSVINAKIADAEALLERKVTDPISIRQNAMETLGLKSSLLGTTKDN 817

QY 251 TISAEDVDSLLALKESQAPIQ 272
 DB 818 TISAEDVDSLLALKESQAPIQ 839

RESULT 4
 C97985
 Hypothetical protein phtD [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: C97985
 R:Hostkins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E.
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A:Title: Genome of the Bacterium Streptococcus pneumoniae. Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11544234
 A:Accession: C97985
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-853 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK99711.1; PID:gl5458514; GSPDB:GN00174
 C:Genetics:
 A:Gene: phtD
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 26.3%; Score 1245; DB 2; Length 853;
 Best Local Similarity 76.0%; Pred. No. 1.1e-53;
 Matches 256; Conservative 5; Mismatches 10; Indels 66; Gaps 6;

QY 2 QITYTDEIQVAKLAGKYTTEDGYIFD-----TSWIKKDSLEAER 42
 DB 517 QITYTDEIQVAKLAGKYTTEDGYIFDPRDITSDGDAYVTPHMTSHWIKKDSLEAER 576

QY 43 AAAQAYAKEKGLTPPSTDHODSGNTEAKAEAIYNRVKAARKVPLDRMPYINQYTYVEKN 102
 DB 577 AAAQAYAKEKGLTPPSTDHODSGNTEAKAEAIYNRVKAARKVPLDRMPYINQYTYVEKN 636

QY 103 GSLIIPSYDHYHNKFEWDEGLYEAPKGYSLDILLATVYKYYE-----PR-----NAS 151
 DB 637 GSLIIPHYDHYHNKFEWDEGLYEAPKGYSLDILLATVYKYYVEHNPHERPHSDNGFGNAS 696

QY 152 DHVRKNK-----AD-----ODSKPDEKDEHSEVTPHPESEKENHAGLNPSAD 196
 DB 697 DHVRKNKQADTNOTKPEEKQTEKEDEHSEVTPHPESEKENHAGLNPSAD 756

QY 197 NLYKPSYDHYHNKFEWDEGLYEAPKGYSLDILLATVYKYYE-----PSIRQNAMETL 235
 DB 757 NLYKPSYDHYHNKFEWDEGLYEAPKGYSLDILLATVYKYYE-----PSIRQNAMETL 816

QY 236 TGLKSSLLGTTKDNNTISAEVDSLLALKESQAPIQ 272
 DB 817 TGLKSSLLGTTKDNNTISAEVDSLLALKESQAPIQ 853

RESULT 5
 B95136
 conserved domain protein SP1174 [imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: B95136
 R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H.
 on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel
 sen, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
 A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: B95136
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-819 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK75283.1; PID:gl4972654; GSPDB:GN00164; TIGR:
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SP1174
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 20.0%; Score 946.5; DB 2; Length 819;
 Best Local Similarity 63.8%; Pred. No. 5e-39; Indels 49; Gaps 7;
 Matches 199; Conservative 23; Mismatches 41;

QY 2 QITYTDEIQVAKLAGKYTTEDGYIFD-----TSWIKKDSLEAER 42
 DB 511 QITYTDEIQVAKLAGKYTTEDGYIFDPRDITSDGDAYVTPHMTSHWIKKDSLEAER 570

QY 43 AAAQAYAKEKGLTPPSTDHODSGNTEAKAEAIYNRVKAARKVPLDRMPYINQYTYVEKN 102
 DB 571 AAAQAYAKEKGLTPPSTDHODSGNTEAKAEAIYNRVKAARKVPLDRMPYINQYTYVEKN 630

QY 103 GSLIIPSYDHYHNKFEWDEGLYEAPKGYSLDILLATVYKYYE-----PR-----NAS 151
 DB 631 GSLIIPHYDHYHNKFEWDEGLYEAPKGYTLEDLLATVYKYYVEHNPHERPHSDNGFGNAS 690

QY 152 DHVRKNK-----ADQSKPDEKDEHSEVTPHPESEKENHAGLNPSADNLYKPSYDTDETEEE 198
 DB 691 DHVRKNKQADTNOTKPEEKQTEKEDEHSEVTPHPESEKENHAGLNPSADNLYKPSYDTDETEEE 750

QY 199 YKPSYDHYHNKFEWDEGLYEAPKGYSLDILLATVYKYYE-----PR-----NAS 235
 DB 751 EEPQVETEKEKLEADLLKQI---DPIIKSNAKETLGLKNNLLGTTQDNNTIMAE 807

QY 256 VDSLLALKESQ 267
 DB 808 AEKLLALKESK 819

RESULT 6
 D98004
 histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: D98004
 R:Hostkins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.

QY	43	AAAQAYAKKGGTTPPSTDHQDSGNTAKGA	EAIYNNRKA	AKKVPIDRDPYNLYQYTV	EYKN	102	
Db	573	VAAQAYTKEGILPPSPDADVKANPTGDS	AAAIYNNR	KGKRIPLVRLPYMYEHT	VEYKN	632	
QY	103	GSLIPSYDHYHNIKFEWDEGLYEA	PKCYSL	EDLLATVKY	YVE-----PR-----	NAS	151
Db	633	GNDIIPKDHYNIKFAWDDHTYKAPNG	YTIEDL	FATIKYVEHPDERPH	SDNGWGNAS	692	
QY	152	DHYRNKKAQDQSKPDRDDEKHDE	VSEPTHP	SESKENHAGLNPSAD	NLYKPSDTDETEE	211	
Db	693	EHV-LGKDBHSDPNKNFRADE--	EPVETPA	EP-----	VPQVETKEVQAQ	736	
QY	212	AEDTTDEAEI-----PCTPSI	RONAMETTLG	LSKSLLLGTYKDN	NTISAEVDSLLALLK	267	
Db	737	LK-----EAEVLLAKVYDSSL	KANATETLAG	LRNNLTLOIMDN	NSIMAEAEKLLALLK	792	
QY	268	PAPIQGPOI	276				
Db	793	PSVSKEKI	801				
RESULT 8							
E98004							
hypothetical protein phtA [imported] - Streptococcus pneumoniae (strain R6)							
C:Species: Streptococcus pneumoniae							
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001							
C:Accession: E98004							
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; De							
y, P.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; M							
e, R.; Sun, P.M.; Winkler, M.E.							
J. Bacteriol. 183, 5709-5717, 2001							
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; J							
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.							
A:Reference number: A97872; MUID:21429245; PMID:11544234							
A:Accession: E98004							
A>Status: preliminary							
A:Molecule type: DNA							
A:Residues: 1-828 <GR>							
A:Cross-references: GB>AE007317; PIDN:AAK99865.1; PID:g15458683; GSPDB:GN01							

C: Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match	13.5%	Score 636;	DB 2;	Length 828;
Best Local Similarity	45.3%	Pred. No. 9.3e-24;		
Matches 140;	Conservative 40;	Mismatches 75;	Indels 54;	Gaps

QY 2 QIYTTDDEIQVAKLAGKYTDEGYFD-----TSHIKDSLSAEAR 42
|| |||:::|::| ||| ||||| | : ||||| : |

Db 539 QIETDEYRIAQADKYTSDGYTFDEHDIISDEGDAYTTPMHGSHHWIGKDSLSDKEK 598
|| |||:::|::| ||| ||||| | : ||||| : |

QY 43 AAQAQAKYKEGLTTPSTDHQDSCGNTAEKAGAEAIYNNRVAAKVKPLDRMPYNLQYTVENK 102
||||| ||||| : ||| ||||| | : ||||| : |

Db 599 VAAQAATKEKGLTTPSPDADVKNPAGTDSAAAIYNNVGEKRIPLVRLPYVYHTVEVKN 658
||||| ||||| : ||| ||||| | : ||||| : |

QY 103 GSLLIIPSDHYHNIRKFEWDEGLYAPKQYSLIEDLLATVKYVVE-----PR-----NAS 151
|| ||| ||||| ||||| ||||| : ||| ||||| ||||| : ||||| : |

Db 659 GNLLIIPKDYHNIRKFAWPDHDTYKAPNGYTTLEDLFIATIKYVVEHPDERPHRSNDGWGNAS 718
|| ||| ||||| ||||| ||||| : ||| ||||| ||||| : ||||| : |

QY 152 DAVRKNKADQSDKPEDKEHDEVPETHPESEDEKENHAGLNPSADNLYKPSDTDETEEE 211
|| || | : : : : ||| ||| ||| : ||| : |

Db 719 ERV-LGKKHSEDPPNKNFKADE--EPVEETPAEPE-----VPQVETEKVEAQ 762
|| ||| ||||| ||||| ||||| : ||||| : |

QY 212 AEDTTDEAIE---PCTPSIRONAMETITGLKSSLLIIGTKNNNTISAEVDSLLALLKESQ 267
||| : ||| : ||| ||||| : ||| : ||||| : |

Db 763 LK---EAEVLLAKYTDSSKANAYETLAGLNNLTQIQMDNNSIWAEEAKLLALLKGSN 818
||| : ||| : ||| ||||| : ||| : ||||| : |

QY 268 PAPIQGPQI 276
|| : : : ||

Db 819 PPSVSKRKI 827
|| : : : ||

A:Reference number: A38455; MUID:91185402; PMID:2010462

A:Accession: A38455

A:Molecule type: DNA

A:Residues: 1-389, TA, 392-724, S', 726-1790 <NA>

A:Cross-references: GB:X34378; MID:94777; PID:CAA38253.1; PID:94778

A>Note: the authors translated the codon ACT for residue 768 as Ile

R:Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.

submitted to the EMBL Data Library, February 1993

A:Description: An integrin analogue in Saccharomyces cerevisiae.

A:Reference number: S30782

A:Accession: S30782

A:Molecule type: DNA

A:Residues: 71-846, E', 848-923, K', 925-1252, I', 1254-1318, V', 1320-1460, S', 1462-1580, S

A:Cross-references: EMBL:L03188

C:Genetics:

A:Gene: SGD:USO1; INT1

A:Cross-references: SGD:S0002216; MIPS:YDI058W

A:Map position: 4L

C:Keywords: coiled coil; transmembrane protein

F:326-342/Domain: transmembrane #status predicted <TM1>

F:394-410/Domain: transmembrane #status predicted <TM2>

F:617-633/Domain: transmembrane #status predicted <TM3>

Query Match 4.3%; Score 201; DB 2; Length 1790;

Best Local Similarity 19.3%; Pred. No. 0.071;

Matches 183; Conservative 132; Mismatches 328; Indels 304; Gaps 40;

QY 67 TEAKGAEIYNVKAQKVPIDRMPYLNQYTVVNGSLIIPSYD-----HYHNKFE 119

DB 650 TTKLQKNTYASHIKQPKK-----DSYFSKDYMDNEDSILTPELDTGLPKRYFTYFIQ 702

QY 120 WFDEGLYAPKGYSLLEDLLATVKYVPEPNASDHVKNKADQSDPKDEKHEV-----174

DB 703 LQENIYRI-----RTALSH-----DPEEPINKISVEEKLOR 737

QY 175 -----SEPHPSDEK-----ENHAGLNPSADNLYKPSDTDEE 207

DB 738 QCTKLKGEITSLOTETESTHNEKTEKLIATNEHKEDEKYQLNSSHSL-KENFSILE 796

QY 208 TE-EAEDTTDAEIPGTPSIQKNAMETITGLKSSLLGCTKNNITSAEVDLSLLALKES 266

DB 797 TELKNVRDSLDE-----MTQLRVLETKDKENQALLEYKSTI-----HK 836

QY 267 QPAPIQGIQGIQPTLPNLSATPSPSPINPCTSHKHEEDG-----YGFDA NRILA 318

DB 837 QEDSIKLEKGLTI-----LSQKKAEDGINKMGKDLFALSREMQAV 879

QY 319 EDESGFVMSHGDS---NHVFFKDLTEE-----QIKAAQHLEVKVTHSHGL---DSL 365

DB 880 EENCNLOKEKDSNVNHOKEKSKLKEDIAAKITEKAINLEEMKIQCNLSKEKEHI 939

QY 366 SSEDYDPCNAEMKDLKKEKAGIKQY---GVKRESIV---VNKEKN-ALITYPSGD 419

DB 940 SKELVEYKSRFOSHNLVAKLEKLSANNYKDMAQENESLIKAVESKNESLSQLSNL 999

QY 420 HHADPIDEHK---PVGIGHSHSNYELF-----PEEGVAKKEGKNVYTGELTNV 467

DB 1000 QNKIDMSQEKENFOIERSIEKNIQAKTISDLQTEKEEIIKSDSK-----DEYESQ 1055

QY 468 VNLLN-----STFNQNTLANGOKRVSFSPPELE--KIGINMLVKLITPDGKVLBK 520

DB 1056 ISLLKLEKLETTANDENVNKISILTKTREETEEALAAAYKLNKLETKLETSE-KALKE 1114

QY 521 V-----SGKVFGEVGNIANFELDQPLPGOTFKY--TIASKDP 558

DB 1115 VNEEHLKEEKIQLEKATETATQQLNSLRANLESLEKHEHDLAQLKRYEQAIANK----1171

QY 559 EYSYDGTFTVPTSLAYKMASQIFYPFHAGDTLYRVNPOF-AVPKGTALVRVFOEFHGN 617

DB 1172 ERQYN-----EELSQNDIEITQOENESIKKKNLEGE 1206

QY 618 AYLENNYKVGKILPKLNQCTTTRAGNKIPVTFMANAYLDNQSTIYVEVPILKENOT 677

DB 1207 V-----KAMKSTSEOSN-----LKKSEIDALNLOIKE---LKKNET 1241

QY 678 DRPSILPOFK-----RNKAQNSKLDEKVE--EPTSKVKEKLESE---717

DB 1242 NEASLESIKSVESVETVKIKELQDCNFKREKEVSELEDKASEDNKSKYLELOKESKI 1301

QY 718 ----TGNSTNSTLEEVPTVDPOEK-----VAKPAESYGMKLENVLFNM 758

DB 1302 KEELDAKTTKIQLEKIKTNLSKAKESELSRLKKTSSERKNAEOLKLN-----1356

QY 759 DGTIELYLPFSGEVIKKNMADFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPE 818

DB 1357 -----EQIKNQK-----PEKERKLLNEGS-STITQEYSKINTLEDELIR 1396

QY 819 APNEKPVTPENSTFNGMLNPEGNVGSDDPMLDPALEAPAVDPVQEKL 865

DB 1397 LQNELKAKKIDNTRSELEKVSLSNDELLE---EKQNTIKSLQDEI 1440

RESULT 13

F95133

Immunoglobulin A1 proteinase [Imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 14-Sep-2001

C:Accession: F95133

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H

non, T.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris

A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: F95133

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-2004 <CUR>

A:Cross-references: GB:A005672; PID:AAK75263.1; PID:g14972632; GSPDB:GN00164; TIGR:

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP1154

C:Superfamily: Streptococcus sanguis IgA-specific metalloendopeptidase

Query Match 4.3%; Score 201; DB 2; Length 2004;

Best Local Similarity 21.0%; Pred. No. 0.083;

Matches 185; Conservative 112; Mismatches 250; Indels 332; Gaps 48;

QY 156 KKAQODSPDEKDE--HDEVSEPHPEDEKENHAGLNPSADNLYKPSDTDETEEEAE 213

DB 51 KYVADSELSEKQQLVYD---IPTVNDDETYTL-----VYKLSNQ-----92

QY 214 DTTDAEIPGTPS-----IRQNAETLGLKSSLLGTDKNT 251

DB 93 ----LAELPTGSKNERQALVAGASLAANGILFAVSKKKVKNKTVLHLVLVAGINGVL 148

QY 252 ISAEVDSLALAK-----ESQAP-----IQGPQIGQPTLPN--NSLA 287

DB 149 VSVHALENHLLNNTDYELTSGEKLPLPKELSGYTYIGYKEGKTSSEVSQKSSVA 208

QY 288 TSPSLPINTPGTSHKHEEDGYGFDA-----RIADESGFVMSHGSHNHYFFKDKIT 341

DB 209 TPT-----KQKVDYNNVTFNFDHPSTVQAIOEQTPVSTKTPTEVQVVEKPPS 256

QY 342 EQIKAAQHLEEVKTSNGLDLSLSEHDYDPCNAEMKDLKKEEKTAGIMKQGVGR 401

DB 257 TELINPR-----KEEKQSDSQQLAEH-----KNLETKKEEKS-----291

QY 402 ESIYVYKKNALITYPSGDHHDAPIDE-----HKPVGIGHSHSNLYPKPEGVAKREG 455

DB 292 -----PKETGV-----NTLNPDQVLSGQINP-----ELLYRETHE-----325

QY 456 NKVYTGEEELTNVNLKNSTFNNQFTLANGOKRVSFSPPELEKLGINN-LVKLIT--512

DB 326 TKIDFQEEI-----QENPDLAEGTVRV-----KQEGKLGKKEIVRIFSVN 366

QY 513 -----PDGKVLKYSGR--VFGE-----GV-----GNIANFELDQPIPL 544
Db 367 KEESRETVSTTAPSPRIVEKGTGKTQVKEQPEGVKQVSGAIVEPAI-QPELP 425
QY 545 GQTKFYIASKDYDEYDGTFTVPTSLAYKMASQTIFYFPACDQTYLVRNPQ-----FAV 600
Db 426 -----EAVVSDKGEVQD-----TLPEAVV-----TDKGET--EVQPESPDTWS 464
QY 601 PKGTDALVRVDEPHGNAYLENNKVEIKR--LPKPL-NQGTTRTAGNKIPVTFMANAY 657
Db 465 DKGPEQVAPLPEKGN-----IQVPEVTEVTKQGPKEK--EEVPV-----507
QY 658 LDNOSTVIVEPILEKNTQDKPSILPQFKRN-----KAQENSKLDEYEEPTSEK-----709
Db 508 -----KPEETFPVNEGTTGTSI--QEAENPVQPAEESTNSEKV-SPDTSKNTGEV 559
QY 710 -----VEKEKLSGTSTGNSISLEEVPTVDPVQEVAKFAESVGMKLENVLF 756
Db 560 SSNPDSSTTSVGEKNGKPEHNDKNSKSEKTEVEVP-VNP-----597
QY 757 NMDGTIELYLPSEGVKKNMADFTGEAPOGNENKPSGKGV---STGTVENOPTENKP- 812
Db 598 -NEGVE-----GTSNQETEPVQPAETQTSNGKIANENGEVSKPSDKPP 645
QY 813 -ADSLPEAPNEKPKVPENS-----TDNGMLNPE---GNVSDPMLDPALEAPAYDPVQEK 864
Db 646 VEESNQPEKNGTATKPSNGTNGTEPEPSNGSTEDVSTESNTSNGNEEIKQE 705
QY 865 LEKTASVGLDLSVIFNMGTIELRLPSGEVKNLSD 903
Db 706 NE-----LDPKKVEEPEKTELEK-----NVSD 728

RESULT 14
S20117

Protein kinase BCK1 (EC 2.7.1.1) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J0906; protein kinase SUK1; protein kinase SSP31; protein YH
C;Species: Saccharomyces cerevisiae
C;Date: 23-Apr-1993 #sequence,Revision 23-Apr-1993 #text,change 24-Sep-1999
C;Accession: S20117; S20298; S22285; S19061; JQ1432; S56872; S30794; JQ1118
Mol. Cell. Biol. 12, 1163-1178, 1992
A;Title: A synthetic lethal screen identifies SUK1, a novel protein kinase homolog impl
A;Reference number: S20117; MUID:92186847; PMID:1545797
A;Accession: S20117
A;Molecule type: DNA
A;Residues: 1-1478 <COS>
A;Cross-references: EMBL:M84389
A;Experimental source: strain S288C
R;Miosga, T.; Boles, E.; Schaff-Gerstenschlaeger, I.; Schmitt, S.; Zimmermann, F.K.
Yeast 10, 1481-1488, 1994
A;Title: Sequence and function analysis of a 9.74 kb fragment of Saccharomyces cerevisia
A;Reference number: S20298; MUID:95176706; PMID:7871887
A;Accession: S20298
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1478 <MIO>
A;Cross-references: EMBL:X77923; MUID:9640004; PIDN:CAA54896.1; PID:9640009
R;Lee, K.S.; Levin, D.E.
Mol. Cell. Biol. 12, 172-182, 1992
A;Title: Dominant mutations in a gene encoding a putative protein kinase (BCK1) bypass t
A;Reference number: S22285; MUID:92107166; PMID:1729597
A;Accession: S22285
A;Molecule type: DNA
A;Residues: 1-58, 'I', 60-1478 <LEE>
A;Cross-references: EMBL:X60227
A;Experimental source: strain EG123
R;Lee, K.S.; Levin, D.E.
submitted to the EMBL Data Library, June 1991
A;Description: An extragenic suppressor of mutations in the S. cerevisiae protein kinase
A;Reference number: S19061
A;Accession: S19061

A;Molecule type: DNA
A;Residues: 1-58, 'I', 60-263, 'P', 265-278, 'I', 280-702, 'S', 707-708, 'KP', 714, 'VTMT', 715
A;Cross-references: EMBL:X60227; MUID:93414; PIDN:CAA42788.1; PID:93415
A;Experimental source: strain EG123
R;Irie, K.; Araki, H.; Oshima, Y.
Gene 108, 139-144, 1991
A;Title: A new protein kinase, SSP31, modulating the SMP3 gene-product involved in pl
A;Reference number: JQ1432; MUID:92104496; PMID:1840547
A;Accession: JQ1432
A;Molecule type: DNA
A;Residues: 149-1478 <IRI>
A;Cross-references: EMBL:D10389; DBJ:D90446
R;Miosga, T.; Schaff-Gerstenschlaeger, I.; Baur, A.; Boles, E.; Chawalatsis, N.; Four
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56855
A;Accession: S56872
A;Molecule type: DNA
A;Residues: 1-1478 <MIR>
A;Cross-references: EMBL:249370; MUID:91008269; PIDN:CAA89389.1; PID:91008270; MIPS:YJ
R;Cusick, M.E.
submitted to the EMBL Data Library, March 1992
A;Reference number: S27437
A;Accession: S30794
A;Molecule type: DNA
A;Residues: 502-959, 'R', 961, 'R', 963-1085, 'V', 1087, 'SLLIHT', 1092-1094, 'RMD', 1101, 'TV'
A;Cross-references: EMBL:M88604; MUID:9172073; PIDN:AAA211179.1; PID:9172074
C;Genetics:
A;Gene: SGD:BCK1; SLK1; SSP31
A;Cross-references: SGD:S0003631; MIPS:YJL095W
A;Map position: 10L
C;Function:
A;Description: phosphotransferase; protein kinase; involved in cell proliferation
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C;Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific protei
F;1173-1440/Domain: protein kinase homology <KIN>
F;1181-1189/Region: protein kinase ATP-binding motif

Query Match 4.28; Score 197; DB 2; Length 1478;
Best Local Similarity 17.74; Pred. No. 0.084;
Matches 183; Conservative 139; Mismatches 307; Indels 404; Gaps 47;
QY 34 KDSLSAERAAOAYAKEGLTTPDST-----HODSGNTEAKAGAEIYNRV-----KAA 82
Db 251 QEDISNRSSTESALSTKS-GPKSTDEKFLHSTSHQKTKSASSLYRFSILRGSSS 309
QY 83 KKVPLDRMPYNLOYTVYEVKNGSLI-----IPSYD-----HYHNKIFE----- 119
Db 310 SNASSAKSPSNIKLSIPARPHSIIESNTLTKSASPSPSPSIFRRHHKSSSESLL 369
QY 120 --WFDGLYEA-----PKGYSLDLLATVYKYYEPRNADSHVKNKADQSKPDEKEH 171
Db 370 NSLFGSGIGEEAPTCKPNQGHSL-----SSENTAKGK-----SKHYETNVS 410
QY 172 DEVSEPHTPSEDEKEN--HAGLNPSADNLKYPSTDTETEEAEEDTTDEARIPCTPSIQ 229
Db 411 SPLQSSILPTSDDKGNLWKKFKRSQIGVSPNVAIVTSQE-----TPSLKS 458
QY 230 NA-----MFTL-----TGLKSLLLGTNDKNTTISAEVDSLLALLKESQ 267
Db 459 NSSTATLTQVADVNIPSPSSPPPIPKTANRSLEVTEDTPKIS-----STTASFETY 514
QY 268 PAPIQGIQGIPTLPNNSLATPSPSLPINCTSHKHH-----EEDGCGFDANR-- 315
Db 515 PDCINPDK-----TVP-----VPVNNQKYSVKNFLLDQKFPYPLKTKGLNDSNNKI 560
QY 316 IIAEDSGFV-----MSH-----GDSNHIF 335
Db 561 LYTKDNVSFVPLNLKSVAKLSFPKESALTGLGINKNVTFTHTDQCDIGAAIPDITLFE 620
QY 336 FKDDL-----TEBOIKAAQKHLEVTKSHNG--LDSLSSEHQDYPGNAKEMKOLD- 383
Db 621 LKSLFLNTSGKIYIKQDKMLQKPKAPLTSENNVPLKSVKSSNRSGTSSLIATSD 680

```

384 -----KIEKIAGIMKQY-----GVKRESIVVNKEK 410
Db 681 VSIYSSSDITSFDEHAGSGRGYQTPSYDYDRVNTNPTEELNYWKE---VLSHEE 737
QY 411 NA--IYPSGDHHDADIDEHK-----PVGIGSHSNYELFKPEGVA---KKGKNVY 459
Db 738 NAKVAVKTSPLKLENLDPKSKLNPITITENESKSFQVLKDEGTIDFNHRESPY 797
QY 460 TGEEL-----INVNLLANSFNQNF'LANGQKRVFS-----PPPELEKL 502
Db 798 TKPELAPKREAPKPPANTSPORTLSTSK-ONKPIRLVRASTRKISKRKSKPLPPOL---L 853
QY 503 GINMLVKLITPDQKLVKYGKVGEGVGNIANFELDQPLPGOTFKYTIASKDYP----- 558
Db 854 SSIPIASSSSPD-----SLTSSYTPAST--HVLIPQPYKGGAND 889
QY 559 -----EVSVDGFTVPTSLAYKM-----ASQTIYPFHAGDTYLRVNPQFAVPGK 603
Db 890 VMRLKLTQDQSTSP-SLKKQKVRNSSTVSTNSIFYS----- 929
QY 604 TDALVYVDFEFGHAYLENNYKVEIKLPIPLKNGTTRTAGNKIPVTFMANAYILDNOST 663
Db 930 -----PSLLKRGNSKRV---VSSTSAADIFEENDIT 958
QY 664 YIVEVPILKENQTKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSSETGNSTS 723
Db 959 FADAPPMFSDSDSDSS-----SSDIIWSKKKTAPETNENKDKSDNSSTHS 1009
QY 724 NSTLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVKKNMADPTGEA 783
Db 1010 DEIFYDSOTQDKMERKMT-----FRPSPEVYQNLKFP---F 1043
QY 784 POGNGENKPSGKSVSTGVENOPTENKPADSI-----PEAPNEKPVKPNST 831
Db 1044 PRAN-LDKPITTEGAS-----PTSPKSLDLSLSPKNVASSRTEPSTPS-RPVPPDSY 1094
QY 832 D-----NGKLN 838
Db 1095 EFTQDGLNGKNP 1107

```

```

RESULT 15
T34418
hypothetical protein F12F3.3--Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Fulton, B.; Wohlmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F12F3.
A:Reference number: Z21521
A:Accession: T34418
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3488 <full>
A:Cross-References: EMBL:U80022; PIDN:AAC2585.1; GSPDB:GN00023; CESP:F12F3.3
A:Experimental source: strain Bristol N2; clone F12F3
C:Genetics:
A:Gene: CESP.F12F3.3
A:Map position: 5
A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

```

```

Query Match 4.28; Score 197; DB 2; Length 3488;
Best Local Similarity 18.78; Pred. No. 0.29;
Matches 108; Conservative 151; Mismatches 433; Indels 234; Gaps 37;

```

```

QY 6 TDDEIQVAKLAGYTTEDGYFTFTSWIKKDSLEAERAAQAYAKEGLPPSTDDHDSG 65
Db 714 TDSKLEAADTTQIETETVDDKS---KKVLKKKTEKSDSFISOKSETPPVPEPTKA 770
QY 66 NTEAKGAEATYNRVK-----AAKVPIDRMPY---NLQTVVEKNGSLI 106
Db 771 ESEAQKIAEV-NKAKQKQVEDDNLKREAEVAAKKIADKLEAEANIKKTAEEVAK-- 827

```

```

Search completed: March 27, 2003, 11:57:55
Job time : 63 secs

```

```

QY 107 IPSYDHYHNIKFEWFEDEGLYEAPKGYSLDGLLATVVKYVVEPRNASDHVR-----NKKA 159
Db 828 -KQEKDQKLETEVYVSKSAEKELEK-OAQIKRAE-----ADAVKKQKELNEKNKL 881
QY 160 DODKSPDEK-----EHDEVSEPHSPDSDEKENHAGLNPSADNLYKSTDTETEETEE 210
Db 882 EAKKSAADKLKLEESAASKVKYSESVKFECKTKRAG-----BKTQVQSEEP 931
QY 211 EAEVTTDAEIPGTPSIRQNAWETLGLKSLLLGTQDNNTIS--AEVDSLALLAKESOP 268
Db 932 TSKTIDTKDVGATEPADETPKKKIHKKTE-----KSDSISQKSATDSKVSQKQKQD 986
QY 269 APIQPOIGQPTLPNNSLATPSPSLPINPGTSHKHEEDGYGFDAIRIIAEDSGFVMSH 328
Db 987 EPT-----KPAVSTQNVTEADK-----SKQKETDEKLKLDAAEIAAKTKQEADEKSK 1034
QY 329 GDSNHYFFKCOLTEBQIIKAAOKHL-----EEVYTSINGLDSLSHSEQDYPGN----- 375
Db 1035 LDAOEKI--KKVSEDDAARKEKELNDKLKLESEIATKASADKLKLEQQAQAKAAVEYA 1092
QY 376 AKEMKDLDKKIE-----EKIAGIMKOYGVKRESIVVNKEKNAIYPSG 418
Db 1093 AKKQKQEKDEQLKLDTEAASKAAAEKLEKQAOQIKKAAGD---AVKKQKE----- 1141
QY 419 DHHADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKVTYGEELTNVNLKKNSTFNN 478
Db 1142 -----LDEKNKLEANKSRAAGLKIEESAASK---QTVQEQALDAQTKARTAEK 1190
QY 479 QNFTLANGQKRVSPFPPELEKKLINMLVLIITPDGKVLK-----VSRKV 525
Db 1191 QT-KLEDEKSTK-----ESESKEVTDE-----KPKKKVYLKKKTEKSDSISQKSETSKT 1239
QY 526 FGEVGNIANFELDQPLPGOTFKYTIASKDYEV-----SYDGTFTVPTS 571
Db 1240 VYESAGPSES-----ETQKVADAAARKQETDEKQKLEAEITAKKSADEKSKLEAE 1289
QY 572 LAYKMASOTIFYPFHAG-----DTYLRVNPQFAPVPGKTDALVRVDFEFGHAYLENNYK 625
Db 1290 SKLKAAAEV-----EAAKKQKQEKQKLDTEAASKAAAEKL-----ELEKQSHIKAAE 1340
QY 626 VGEIKLPIPLKNGTTRTAGNKIPVTFMANA---YLDNQSTYIVVPILEKENOTDKPSI 682
Db 1341 VDAVK-----KQKELEEKQKLESEAAATKKADEKLEOKKAAEATLIEIOKEQK--- 1393
QY 683 LPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSSETGNSTNSTLEEVPTVDPQVKVAK 742
Db 1394 -----LAQEQSRLEDEAKSAEKQKLESETKSKQTEAPKESVDEKPKKKVLKKKTEK 1446
QY 743 FAESYGMKLENVLFNMDGTIELYLPSEGVKKNM-----ADTGAPOQNGEN 790
Db 1447 SDSSISQKSASAKSTVDAAEATLESDFNLVEKKTQVKQEQSPDESTSATIKRDPQAKTEBI 1506
QY 791 KPSGKSVSTGVENOPTENKPADSLPEAPNEKPVKPNSTDNGLNPNNGVSGDPMLOP 850
Db 1507 SKQDDGDEKKTITDGGPP--KPEDS--EAPPKRVVKKTKQKSDSVASDASLADVSKLSD 1562
QY 851 ALEAPAPVDPQVKLEKFTASTYGLDLSVIFNMDGTIELRLPSGEV 896
Db 1563 DVEEKPKKKVLAKKTEK-----SDSVISETSSVDITIKPESVEI 1600

```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:53:12 ; Search time 27 seconds
(without alignments)
1391.761 Million cell updates/sec

Title: US-09-884-465A-332

Perfect score: 4728

Sequence: 1 MQITYTDEIQVAKLAKYT.....IELRLPSGEVKKNLSDFFIA 906

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	204	4.3	1790	1 US01_YEAST	P25386 saccharomyc
2	197	4.2	1478	1 BCK1_YEAST	Q01389 saccharomyc
3	194.5	4.1	1164	1 BAG_STRAG	P27951 streptococc
4	189.5	4.0	2748	1 NDM1_YEAST	Q00402 saccharomyc
5	189	4.0	1637	1 MSP1_PLAFW	P80544 staphylococ
6	186	3.9	1658	1 YMP7_YEAST	Q03661 saccharomyc
7	183	3.9	1636	1 BUD3_YEAST	P25558 saccharomyc
8	181.5	3.8	1139	1 HWM1_MYCGE	Q49413 mycoplasma
9	180.5	3.8	1183	1 CNA_STAAU	Q53654 staphylococ
10	180	3.8	1639	1 MSP1_PLAFW	P04933 plasmodium
11	176.5	3.7	2195	1 SC16_YEAST	P48415 saccharomyc
12	176	3.7	2738	1 PCGV_RAT	Q98rb4 rattus norv
13	174	3.7	3924	1 ANK2_HUMAN	Q01484 homo sapien
14	173.5	3.7	710	1 L778_ARATH	Q06738 arabidopsis
15	172.5	3.6	1701	1 MSP1_PLAFW	P08569 plasmodium
16	171.5	3.6	825	1 SW13_YEAST	P32591 saccharomyc
17	171	3.6	1240	1 YN11_YEAST	P53935 saccharomyc
18	171	3.6	1630	1 MSP1_PLAFK	P04932 plasmodium
19	170.5	3.6	1545	1 IG43_HAEIN	P45385 haemophilus
20	170.5	3.6	1701	1 MSP1_PLAFW	P38119 plasmodium
21	169.5	3.6	2805	1 MAPA_HUMAN	P78559 homo sapien
22	169	3.6	639	1 ACTA_LISMO	P33379 listeria mo
23	168.5	3.6	3135	1 S230_PLAFO	Q08372 plasmodium
24	168	3.6	1612	1 T22B_MOUSE	Q64511 mus musculu
25	167.5	3.5	1726	1 MSP1_PLAFW	P04934 plasmodium
26	167.5	3.5	1726	1 MSP1_PLAFW	P50495 plasmodium
27	166.5	3.5	1233	1 YF16_YEAST	P43597 saccharomyc
28	166.5	3.5	1612	1 T22B_CRIO	Q64399 cricetus
29	166	3.5	1616	1 P200_MYCGE	Q49429 mycoplasma
30	165.5	3.5	1181	1 SC42_STRPY	P58099 streptococc
31	165	3.5	1849	1 IG44_HAEIN	P45386 haemophilus
32	164.5	3.5	3381	1 PCGV_BOVIN	P81282 bos taurus
33	163.5	3.5	2869	1 RBP1_PLAVB	Q00798 plasmodium

34	163	3.4	1466	1 SPA2_YEAST	P23201 saccharomyc
35	163	3.4	2774	1 MAPA_RAT	P34926 rattus norv
36	162.5	3.4	1766	1 SFLL_YEAST	P20134 saccharomyc
37	161.5	3.4	1781	1 AKAC_HUMAN	Q02952 homo sapien
38	161.5	3.4	1251	1 RBP2_PLAVB	Q00799 plasmodium
39	161	3.4	1167	1 SCAL_STRPY	P15926 streptococc
40	161	3.4	3122	1 POZ_MOUSE	Q61493 mus musculu
41	160.5	3.4	1095	1 NEB1_RAT	Q35867 rattus norv
42	159.5	3.4	1381	1 YBE7_YEAST	P34216 saccharomyc
43	159.5	3.4	1664	1 INT1_CANAL	P33705 candida alb
44	159	3.4	1007	1 RGAL_YEAST	P39083 saccharomyc
45	159	3.4	1189	1 YJH6_YEAST	P47035 saccharomyc

ALIGNMENTS

RESULT 1
US01_YEAST STANDARD; PRT: 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein US01.
GN US01 OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91183402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, us01, is required for intracellular
protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC '- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
COMPLEX.
CC '- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH INTRACELLULAR
MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
ER AND THE GOLGI COMPLEX.
CC '- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC '- SIMILARITY: BELONGS TO THE VDP/US01/YBL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; X54378; CAA38253.1; -;
CC EMBL; L03188; AAB00143.1; -;
CC EMBL; U53668; AAB66659.1; -;
CC PIR; A38455; A38455.
CC SGO; S0002216; US01.
CC InterPro; IPR002017; Spectrin.
CC Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.

Query Match 4.39; Score 204; DB 1; Length 1790;
Best Local Similarity 19.79; Pred. No. 0.033;
Matches 182; Conservative 141; Mismatches 342; Indels 260; Gaps 41;

DB 67 TEAKGAEAIYNRVKAARKVPLDRPYNLYQYVEKNGSLIIPSYD-----HYHNKFE 119
DB 650 TKTGLKONVARIKQFK-----DSYFSKVDNEDSILPELDEGLPKVYESTYFIQ 702
QY 120 WFDGLYKAPKYSL-----EDLATYKYVPEPNASDHVKNKADODS-KPEDKEHDEVS 175
DB 703 LFENIYRIRIALSHDPEEPISKIS-PEEVEKLQROCTKLKGETTSLOTETSTHENLT 761
QY 176 EP-----THESDEKENHAGLNSADNLYKPSDTETE-EAEDYTDAAEIPGTFSIR 228
DB 762 EKLIAT/WEHKEDEK---YQILNSSHSL-KENFSILETLKNVYRSLDE-----808
QY 229 QNAMETLGLASSLLGTYKONNTTSAEYDLSLALLKESQAPIQGQPTLPNNSLAT 288
DB 809 -----WQLRDVLTKEKQNTALLEKYSI-----HQEDSIKLEKLETI-----851
QY 289 PSPPLNPGTSHEKHERG-----YGDANRIADDESGFVMSHGD-----NHVFFK 337
DB 852 -----LSQKKAEDGINKMGKDLFALSREMQAEECNKQLQKEKDSNVNHOQET 901
QY 338 KDLTEE-----QIKAAOKHLEKVTSHGL-----DSLSEHEDYVGNKEMKDLDDKIE 387
DB 902 KSLKEDI/IAKTEIKAINENLEEMKIQCNLSKEKEHLSKELVETKSRFQSHDNLVAKLT 961
QY 388 EKIAGIKQY---GVKRESIV---VNKERN-ALIYPSGDHHDPTIDEK---PVGIGHSH 438
DB 962 ERLKSLANNYKDMAENSLIKAVEESNESSIQLSNLQKIDMSQKENFQIERGSIE 1021
QY 439 SNYELFK-----PEGVAKREGNKVYGEELTNVNNLKN-----STFNQNFILA 484
DB 1022 KNIEQAKRTISDLEQTEKEEIIISKSDSR-----DEYESQISLLKELTATTANDENVNKI 1077
QY 485 NGOKRVSFSPPELE--KKGINMLVLKITPDGKVLKV-----S 522
DB 1078 SELTYRELELAELAYNKNKNELETKLETSE-KALKEVKEENEHLKEEKTOLEKEATET 1136
QY 523 GRVFGEGVGNIANFELDQYLPFGQTFKY--TASQDYEVSYDGTFTTPTSLAYKMASQT 580
DB 1137 KQOLNSLRANLESLEKEHEDLAQQLKVEEQIANK---EROYN-----1176
QY 581 IYPPHAGDTYLRVNPQF-AVPKGTDALVRVTFDEPHGNAYLENKYGEIKLPKLNQ 639
DB 1177 -----EISQLNDEITSTQOENESIKKKNDLEGEV-----KAMKS 1212
QY 640 TIRTAGNIPVTMANAYLDNOSTIVVPILEKNQNDKPSILPQFK-----687
DB 1213 TSEOSN-----LKSIDALNLOIKE---LKKANETNEASLSLESIKSVSETVKIKEL 1263
QY 688 -----BNKAQENSKLDERVE--EPKTSERKEKLESE-----TGNSTNSNSTLEEVPTVD 734
DB 1264 QDECNFKEKEVSELEDKLAKSDKNSKYLELQKESEKIKEELDANTTELKIOLEKITNLS 1323
QY 735 PVOEK-----VAKFAESYGMKLENVLFNMDGTIELYLPGEVVIKKNADFT 780

DB 1324 KAKESELSRLKKTSSERKNAEQLEKLN-----EIQKNQA---1364
QY 781 GEAPOGNGENKPKENGKSVCTGVENQPTENKPADSLPEAPNEKPVDPENSTGMLNPEG 840
DB 1365 -----FEKERLLNEGS-STTQSEYSEKINTLELIRLQNELKAKEIDNTRSELEKY 1418
QY 841 NVGSDPMDLPALPEAPAVDPVQEK 865
DB 1419 SLNDELLE---EKQNTIKSLQDEI 1440

RESULT 2
BCK1_YEAST
ID BCK1_YEAST STANDARD; PRT; 1478 AA.
AC Q01389; P32894;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Serine/threonine protein kinase BCK1/SLK1/SSP31 (EC 2.7.-.-).
GN BCK1 OR SLK1 OR SSP31 OR LAS3 OR YJL095W OR J0906.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92186847; PubMed-1545797;
RA Costigan C., Gehring S., Snyder M.;
RT "A synthetic lethal screen identifies SLK1, a novel protein kinase
RT homolog implicated in yeast cell morphogenesis and cell growth.";
RL Mol. Cell. Biol. 12:1162-1178(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92104496; PubMed-1840547;
RA Irie K., Araki H., Oshima Y.;
RT "A new protein kinase, SSP31, modulating the SMP3 gene-product
RT involved in plasmid maintenance in Saccharomyces cerevisiae";
RL Gene 108:139-144(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-EG123;
RA MEDLINE-92107166; PubMed-1729597;
RA Lee K.S., Levin D.E.;
RT "Dominant mutations in a gene encoding a putative protein kinase
RT (BCK1) bypass the requirement for a Saccharomyces cerevisiae protein
RT kinase C homolog";
RL Mol. Cell. Biol. 12:172-182(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C;
RA MEDLINE-95176706; PubMed-7871887;
RA Miosga T., Boles E., Schaaff-Gerstenschlaeger I., Schmitt S.,
RA Zimmermann F.K.;
RT "Sequence and function analysis of a 9.74 kb fragment of
RT Saccharomyces cerevisiae chromosome X including the BCK1 gene";
RL Yeast 10:1481-1488(1994).
RN [5]
RP SEQUENCE OF 602-1104 FROM N.A.
RA Cusick M.E.;
RL Submitted (XX-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SERINE/THREONINE PROTEIN KINASE INVOLVED IN A SIGNAL
CC TRANSDUCTION PATHWAY THAT PLAY A ROLE IN YEAST CELL MORPHOGENESIS
CC AND CELL GROWTH. THIS PATHWAY SEEMS TO STARTS BY SMP3; THEN
CC INVOLVE THE KINASE PKC1 THAT MAY ACT ON THIS KINASE. BCK1 PROBABLY
CC PHOSPHORYLATES MKL1 AND MKK2 WHICH THEMSELVES PHOSPHORYLATE THE
CC MPK1 KINASE.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE KINASE KINASE SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

STRAIN-LA239;
MEDLINE-91312121; PubMed-1857207;
Jarlstrom P.G., Chatawal G.S., Timmis K.N.;
"The Iga-binding beta antigen of the c protein complex of Group B streptococci: sequence determination of its gene and detection of two binding regions.";
Mol. Microbiol. 5:843-849(1991).
[2]
IDENTIFICATION OF IG-LIKE DOMAIN.
MEDLINE-37035265; PubMed-888921;
Bateman A., Eddy S.R., Chothia C.;
"Members of the immunoglobulin superfamily in bacteria.";
Protein Sci. 5:1939-1942(1996).
-1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (Potential).
-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
EMBL; X59771; CAA42442.1; --
PIR; S15330; FCSOAG.
InterPro; IPR004829; Surface_antigen.
InterPro; IPR001899; Gram_pos_anchor.
InterPro; IPR003599; Ig.
Pfam; PF00746; Gram_pos_anchor; 1.
ProDom; PD153432; Surface_antigen; 2.
SMART; SM00409; IG; 1.
TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
TIGRFAMs; TIGR01168; YSRK_signal; 1.
PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
Cell wall; Peptidoglycan-anchor; Receptor; Repeat; Signal;
Immunoglobulin domain.
SIGNAL 1 37
CHAIN 38 1135 IGA FC RECEPTOR.
PROPEP 1136 1164 REMOVED BY SORTASE (POTENTIAL).
DOMAIN 434 534 IG-LIKE DOMAIN.
DOMAIN 199 438 IGA-BINDING (POTENTIAL).
DOMAIN 439 826 IGA-BINDING (POTENTIAL).
DOMAIN 827 945 PRO-RICH REPEATS.
SITE 1132 1136 LPXTG SORTING SIGNAL (POTENTIAL).
MOD_RES 1135 1135 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SEQUENCE 1164 AA; 131051 MW; 65DE94AF720A5474 CRC64;
Query Match 4.1%; Score 194.5; DB 1; Length 1164;
Best Local Similarity 19.4%; Pred. No. 0.052;
Matches 166; Conservative 130; Mismatches 316; Indels 245; Gaps 40;
QY 166 DDKEHDEVESETHP---ESDKEHAGLNPSADNLYKPSDTETEEAEYTTDEAIP 222
DB 43 DSVKTEVAAPKATLGLKSL---LGTGDNNTISAEV-----DSILALLKESQAPIQG 273
QY 223 GTPSTRONAKMETLGLKSL---LGTGDNNTISAEV-----DSILALLKESQAPIQG 273
DB 103 DPGKREKQLQWKNLKNVDVNTLSHEQKNEFKYKIDETNDSALLEL-----151
QY 274 POIGQPTLPNNSLATPSPSLPINTGTSHEKHEEDGYGFDPANRIAEDESGFVNSHGDG---331
DB 152 -----ENQFNETNRLHLK---QHEVEKDKKA-KQKTLKQSDTKVDSLNDKREL 198
QY 332 NH-----YFFKDLTEQIKAKHLEVTSHNGLDLSLSSHEQDPGNKEMKDLK 384
DB 199 NHOKSOVKASQKGTNEDKDSMLKIEDIRKQAQADKEDAE-----VKVREELGK 252
QY 385 KTEKTAGIWK---QYGVKRESVVAKENAIYPSGDHHDPI-----426
DB 253 LPSSTAGLDQELQEHVKET---SSEENT---QKVDHEYANSLQNAKLSLELDKATT 306

QY 427 DEHKPVGIGHSHSNYELFKPEGCVAKKEGKVV--TGEELTNVNVLLKSNFTNQNFTLA 484
DB 307 NEQATQVKQNLAKLKEIPLIKETNVKLYKAMSESLEQVEKELKHNSEANLEDLVA 366
QY 485 NGQKRVSFPPPLEKLGINMLVKLITPDGKLVLEKSVGVFGEVGNL-ANFELDQPYL 543
DB 367 KSEIYR-----EYEGKL--NOSKNL--PELKOLEEASHKLAQVVEDFRKFKTSEQVT 417
QY 544 PGQTFKYTIASKD-----YPE--VSYDG---TFTVPTSLAYKMASQTIFFPFHAG 588
DB 418 PKKRVRLAANENNOQKIELTVSPENITVEGEDVKFTVA---KSDSKT---TLDIFS 470
QY 589 DTYLRVNPQPAVPGKTDALVRVDFEFGHGNAYLENNYKVGSEIKLPIKLNQGTTRT-----643
DB 471 DLLTKYNPSVS-----DRISTNYKTNT---DNHKAIEITIKNLKLNESOTVTLKAKD 519
QY 644 -AGNKIPVTMANAYLDNQSTIYVEVPILEKENOTDKPSILPOFKRKAQENSKLDEKV- 701
DB 520 DSGNVVEKTF-----TITVQKKEK-----QVPTPEQKDSKTEKVP 557
QY 702 EEPRTSEKVEKELSETGN-----STSNSTLEEPTV-----DPVOE 738
DB 558 QEPKSNQKQLOELIKSAQOQLEKLEKAIKELMEQPEIPSNPEVGIQKSIWESQKEPIQE 617
QY 739 KVAKF-----AESYGMKLENVLFNMDGTIELYLPSEGVKKNMADFTGEAPQG 786
DB 618 AITSFKKIIGDSSSKYTYTEHYFNKYSDFMNYQLHAQM-----EMLTRKYVQVNNKYPDN 672
QY 787 NGENKPSENGKVS-----GTVENOPE-----PADSLPEAPNEKP 824
DB 673 AEIKKIPESDMKRTKEDNYGSLNDALKGFPEKFLYFPFNKIKQIVDDDKKVEQDOPAP 732
QY 825 VKPENSTDNGLNPEGNVGDPMDDPALEA-----PAVDPVQEKLEKFTASYGLG 875
DB 733 I-PENSE-----MDQAKERAKIAVSKYMSKVLGDVGHQLOKKNNSKIVD 775
QY 876 L-----DSVIFNMD 884
DB 776 LFELEAIKQOTIFDID 792
RESULT 4
ID NUM1_YEAST STANDARD; PRT; 2748 AA.
AC Q00402; (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nuclear migration protein NUM1.
GN NUM1 OR YDR150W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 28383 / FL100;
RX MEDLINE-92079907; PubMed-1745235;
RA Kormanec J., Schaeff-Gerstenschlaeger I., Zimmermann F.K.,
RA Perecko D., Kuentzel H.;
RT "Nuclear migration in Saccharomyces cerevisiae is controlled by the highly repetitive 313 kDa NUM1 protein.";
RL Mol. Gen. Genet. 230:277-287(1991).
CC -1- FUNCTION: CONTROLS NUCLEAR MIGRATION. NUM1 SPECIFICALLY CONTROLS THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR ENVELOPE.
CC -1- MISCELLANEOUS: ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT CONSENSUS (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH FLANKING DOMAINS OF THE TANDDEM REPEATS.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X61236; CAA43554.1; --
 DR PIR; S19052; S19052.
 DR SGD; S0002557; N0M1.
 DR InterPro; IPR001849; PH.
 DR Pfam; PF00169; PH: 1.
 DR SMART; SM00233; PH: 1.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 KW Repeat.
 DR DOMAIN 593 1384 12.5 X TANDEM REPEATS.
 FT REPEAT 593 656 1.
 FT REPEAT 657 727 2.
 FT REPEAT 728 798 3.
 FT REPEAT 799 862 4.
 FT REPEAT 863 926 5.
 FT REPEAT 927 990 6.
 FT REPEAT 991 1054 7.
 FT REPEAT 1055 1118 8.
 FT REPEAT 1119 1182 9.
 FT REPEAT 1183 1246 10.
 FT REPEAT 1247 1310 11.
 FT REPEAT 1311 1374 12.
 FT REPEAT 1375 1384 13 (INCOMPLETE).
 FT DOMAIN 2573 2683 PH.
 SQ SEQUENCE 2748 AA; 313202 MW; B2FBD67C9F6211AE CRC64;

Query Match 4.0%; Score 189.5; DB 1; Length 2748;
 Best Local Similarity 19.1%; Pred. No. 0.28;
 Matches 196; Conservative 152; Mismatches 339; Indels 341; Gaps 45;

QY 7 DEIOVAKLAGVT-----TEDGY-----IFDTSWIKKDSLEAEARAAQAY 49
 DB 373 DDEMTKEVAENLNTALPNDONYKESLESHEIKYLEASGYKVLPLEFENL----- 426
 QY 50 KEGLTPTPTDHDGNTGAEAIYRNVAAKYPLDRMPYNL---QYVEY-----K 101
 DB 427 -NESLNSPVNY-----LKEQLQAKKIPIDQSTFNLLKEPTIDFLPLTSK 472
 QY 102 NGLIIPSTDHYNIKPEWDEGLYAPGYSLE-----DLLATVKYVVERNAS 151
 DB 473 IDCLITPTKDY-----NDLPESVKNPSTQMKKCLEAKNDLQSNICKWLEERNGC 522
 QY 152 DHVRK-----NKADQDSK---PDDEKHEDEVSEPTHPESDEKENHAGLNPSADNLYKP 201
 DB 523 KWLNDLYSWNNKITPQKYLSDRAKEYDQVL-----IDTKALEGLKNPTIDFLREK 576
 QY 202 STDTEETEEAEADTDEAIEPGPISRONAMEPLTGLKSLGLTGDNDNTISAEVDSLLA 261
 DB 577 ASASDVLKLLKEDYV-----SPSL-EYLVEHAKATNHHLLSDSAYEDLVKCKENPDME 628
 QY 262 LKESOPATIQGPIQGPILPNNSLATPSPSPINPGTSHKEHEEDGYGDFDNRITAEDE 321
 DB 629 FLKE-----KSAKLGHTVVSNEAYS-----ELEKLEQPS 658
 QY 322 SGFVMSHGD-SNHYFFKDLTDEQIKAAOK-HLEEVKTSNGLDLSLSHEDDYPGNAKEM 379
 DB 659 LYLVEHAKATNHHLLSDSAYEDLVKCKENPDMEFLKEKSAKLGHTVVSNEAYSELQRY 718
 QY 380 KOLDKIEEKIAGIMQYGVKRESIVVNNKNAIIPSGDHHADPIDEHKPVGIGHSHS 439
 DB 719 SELEKEVEQPSLAYVE-----HAKATDHL-----LSDS 748
 QY 440 NYE-LKPEEG-----VAKKEGKNVYTGELTNVNNLLKSTNNQNTFLANGKRV 490
 DB 749 AYEDLVKCKENPDVEFLKEKSAKLGHTVVSSEYSEL-----ORKY 789

QY 491 SFSPPPELEKKGIMLVKLI-----TPDGKVLKVSQKVF 526
 DB 790 S-----ELEKEVEQPSLAYLYVEHAKATDHHLLSDSAYEELVKCKENPDMEFLKEKSAKLG 844
 QY 527 GEGYNTIANFEL-----DQYLPQGTQFKYTIASKDYPEVSY----- 562
 DB 845 HTVVSNEAYSELEKLEQPSLAYLYVEHAKATDHHLLSDSAYEDLVKCKENSDVEFLKEKS 904
 QY 563 -DGTFVTPTSLEYKMASQTIFFP-----FHAGDTYL-----RVNPOFAPV 601
 DB 905 AKLGHTVVSNEAYSELEKLEQPSLAYLYVEHAKATDHHLLSDSAYEDLVKCKENPDMEFL 964
 QY 602 KGTDALVRVDFEFGNAYLENN-YKVGIEIKLPIPKLNQGTTRTAGNKIPYVTFMANAYLDN 660
 DB 965 KEKSAKL-----GHTVVSNEAYSELEKLEQPSLAYLYVEHAKA-----TNHHLLS 1009
 QY 661 QSTYIVVEPILEKENQTKPSILPOFKRKA-----QENSKLDEKVEEPTSEKV 710
 DB 1010 DSAY--EDLVKCKEN---PDM--EFLKEKSAKLGHTVVSNEAYSELEKLEQPSLEYLV 1061
 QY 711 EKEKLSGTNSTNSTLEEV-----PTVDPVQEKVAKFA-----ESYGMKLENVLFNM 758
 DB 1062 EHAK--ATNHHLLSDSAYEELVKCKENPDVEFLKEKSAKLGHTVVSNEAYS-ELEK----- 1114
 QY 759 DGTIELYLPSEGVIKKN-----MADFTGEAPOGNGENKPSNGKYSTGTCTVENOPTEN 810
 DB 1115 ---KLEQPSLEYLYVEHAKATNHHLLSDSAYEELVKCKENPDVEFLKEKSAKLGHTVVS 1170
 QY 811 KPADSLPEAPNEKP-----VKPENSTNGMLNPEGVNGSDPMLDPALE-----EAPAVD 859
 DB 1171 EAYSEL-EKKLEQPSLAYLYVEHAKATDHHLLS-----DSAYEDLVKCKENPDVE 1218
 QY 860 PVQEKLEK 867
 DB 1219 FLKEKSAK 1226

RESULT 5
 MRSP_STRAU STANDARD; PRT; 1637 AA.
 AC P80544; Q9ZF62;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Methicillin-resistant surface protein precursor.
 GN PLS.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate 1061;
 RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;
 RT "Pis, a large repeat-rich surface protein of methicillin resistant
 RT Staphylococcus aureus";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 624-628; 676-682; 938-948; 1156-1168; 1176-1185;
 RP 1199-1205 AND 1217-1224.
 RC STRAIN=Isolate 1061;
 RX MEDLINE=96370743; PubMed=8665912;
 RA Hilden P., Savolainen K., Tynnelae J., Vuontola M., Kuusela P.;
 RT "Purification and characterization of a plasmin-sensitive surface
 RT protein of Staphylococcus aureus";
 RL Eur. J. Biochem. 236:904-910(1996).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (Potential).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF115379; AAD09131.1; -
 CC InterPro; IPR001899; Gram_pos_anchor.
 CC Pfam; PF00746; Gram_pos_anchor; 1.
 DR TIGRfam; TIGR01167; LPYTG_anchor; 1.
 DR TIGRfam; TIGR01168; ISIRK_signal; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 KW Antibiotic resistance; Glycoprotein; Cell wall; Peptidoglycan-anchor;
 KW Repeat; Signal.
 FT SIGNAL 1 48 POTENTIAL.
 FT CHAIN 49 1601 METHICILLIN-RESISTANT SURFACE PROTEIN.
 FT PROPEP 1602 1637 REMOVED BY SORTASE (POTENTIAL).
 FT DOMAIN 1301 1584 141 x 2 AA TANDEN REPEATS OF D-[SAG].
 FT SITE 1598 1602 LPYTG SORTING SIGNAL (POTENTIAL).
 FT MOD_RES 1601 1601 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 SQ SEQUENCE 1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;

Query Match 4.0%; Score 189; DB 1; Length 1637;
 Best Local Similarity 19.8%; Pred. No. 0.15;
 Matches 208; Conservative 118; Mismatches 415; Indels 308; Gaps 47;

QY 6 TDEIQVAKLAGYTTEDGYIFDTSWIKKDSLEAERAAQAYAKEGLTPSTDQDQSG 65
 DB 124 TTEAPKAECTDRTVEAPKAEET--DRATTEAPKAEETDKATEA---PKTEBDKA 178
 QY 66 NTE-----AKGAAIYRVKAAKVPDLRPNYQVTVVEKNGSLIIPSYDHYHNK 117
 DB 179 TTEAPAAEETSKAATEAPKAEETSKAATEAP---KAEETKATEAPTEETDKYE 235
 QY 118 FEWFDEGLYAPAGYSLEDLLATKYVYVPRNADVRKADQDSKDEDEHDEVSPP 177
 DB 236 TE-----EAPKAE-----TSKATEAPKAEET--NKVETEAPAEETNAATEE 280
 QY 178 THP-ESDEKENHAGLNPS-----ADNLYKPSDTETEERAE----- 214
 DB 281 TPAVEDTNKASNSNAQPSERTQVVDVTVAKDLYKKSEVTEAEKAEIKVPLPRDISNLSN 340
 QY 215 -----TTQDAEIPG--PSTRONAMETLGLKSL-- --LGTKDN 250
 DB 341 EEIKKIALSVLEKATANKENAPRATRSVSSNARTNINYSATLRAAQAQDTVYKGTG 400
 QY 251 TISAEVDSLALAKESQAPIQGQPTLPNNSLATPSP-----SLPI 295
 DB 401 NPTAGHDIIHKYKEE--PPNEGTLAFNTNPNPTGTGALSEYNDKIDFNKDTITVPV 458
 QY 296 NPGTSHEKHEEDGYGTANRIIAD--ESGFVMSHSDSNHYFFKDLTEEQIKAAQKHL 352
 DB 459 ANNNQGNNTGADNGFMFTQGNQDFLNQGILDRKGMANASGFKIDTAYNNVNGKVDKL 518
 QY 353 BEVTSNGLDLSLSHQDYPGNAKEMKLDKIEEKIAGIMQYGVKRESIVVNEK-- 410
 DB 519 DADYT--NNLSQIAKAVGVGTFKNGAD-----GVTVQGV--QNALNTKDKPV 563
 QY 411 NAIYPSGDHHDADIDEHPPVIGSHS-----NYELPPEGVAKKEKNKYVTE 462
 DB 564 NKIIYADNTNHLD-----GPHGQRLNDVVLNYD-----AATSTTATYAGK 606
 QY 463 ELTVNVLNLSNFNNQFTLANGQRVSFS---FPPELEKKLGINMLVKLITPDGKVL 519
 DB 607 TWKATDGLDGSQKYNFLITSSHQNRYSNGIMRTNLEG-----VTTITPDQADLL 659
 QY 520 KVSQKVEGEGVGNIANPELQDPLPGQTFYTIASKDYPEV-----SYDG--TFVTPTSLA 573
 DB 660 DV--EVTQKPIPHKTIREFDTLEPGS-----PDVIVKQGEDGKTTTPT--- 703
 QY 574 YKMASOTIYFPFHAGDYLRVNFQAVPKG--TDALVNVF-----DEFHGNAYLENN 623
 DB 704 -KVDPDT-----GDVVERGEPTTEVTKPNVDEIVHFTPEEVPQGHKDEFPN----- 749
 QY 624 YKVEIKLPIPLKNGQTRTAGNKIPVTFMANAYLDNQSYIIVEPILK-----E 674

DB 750 -----LPI-----DGTEEVPG-----KPGIKNPETGEVVTVPDDVTKHGPAGE 789
 QY 675 NQTDKPSILPQFRNKAQENSKLDKVEEPTKSEKKEKLSGTGNSNSTLEEV---- 730
 DB 790 PEVTKKEIPPEKKREFNPDLKPGEEKV---TQGGTGEKTTTPTTINPLTGERKGE 845
 QY 731 PTVDVPQOEKVAEASVGMKLE-----NVLFNMDGTIELYLP-----SGEVIK 774
 DB 846 PTVEVTEPDEITQFGEVEPQGHKDEFPNLPIDGTEE--VPGKPGIKNPETGEVVT 903
 QY 775 NMADFTGEAPGN-----GENKPSNGKVSCTGV----- 803
 DB 904 PVDDVTKHGPAGEPEVTKKEIPPEKKREFNPDLKPGEEKVTOGGTGEKTTTPTTINP 963
 QY 804 -----ENQPT-----ENKPADSLPEAPNEK-----PVKPNSTDN-----GMLNP 838
 DB 964 LTGEKVEGEPTTEVTEPDEITQFGEVEPQGHKDEFPNLPIDGTEEVPQGHKGNP 1023
 QY 839 E-GNVGSDPMLDPALEAPAVDPVQEKLE 866
 DB 1024 ETGEVVTVPDDVTKHGPAGEPEVTKKEE 1052

RESULT 6

YMF7_YEAST STANDARD; PRT; 1658 AA.
 ID YMF7_YEAST
 AC Q03661; Q04988;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 187.1 kDa protein in GUAL-ERG8 intergenic region.
 GN YMR219W OR YMR261.13 OR YMR959.01.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OK NCBI_Taxid:4932;
 RN [1]
 RP SEQUENCE OF 1-711 FROM N.A.
 RC STRAIN-S288c / AB972;
 RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
 RA Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 608-1648 FROM N.A.
 RC STRAIN-S288c / AB972;
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; Z49809; CAA89934.1; -
 DR EMBL; Z49939; CAA90190.1; -
 DR SGD; S0004832; YMR219W.
 KW Hypothetical protein.
 SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 3.9%; Score 186; DB 1; Length 1658;
 Best Local Similarity 19.1%; Pred. No. 0.21;
 Matches 217; Conservative 164; Mismatches 389; Indels 366; Gaps 53;

QY 6 TDEIQVAKLACK---YTTEDGYIFDTSWIKKDSLEAERAAQAYAKEGLTPSTD- 60
 DB 583 SEEQIPKDTGENTNNKTDGRDLSSSVIEIEVKVSEKKLDG-----STRELPLSLDT 638
 QY 61 -----HQDSGNTGAEAGAAIYRVKAAKVPDLRPNYQVTVVEKNGSLIIPSYD 112

Db 639 TINSSGLNEDSIYSLDADAISENL---TDVPLMEIKTTPKYEIVISESYSTSYED 695
QY 113 -----YHNKFEWDEGLYAPKGYSLDLELATVKKYVEB----- 147
Db 696 NTVAMPQVETSPMNDPNSLND--YB-----KKHLLKSLALAPAFYKKADEFVE 749
QY 148 -----RNASHDRVKKKADQSK--PDDEKHEHVESEPTHPSEDEKENHAGLNPSADN 197
Db 750 AGVTKSCLTSTSGHTNIPHTSKETQVSLDSESTENVTENENTGDNKNQKNPVGVA 809
QY 198 LKVPSTDTETEBAEDTDE-----AETPGPSIRONAMETUTGLKSSLLGKTGN 249
Db 810 -----STDKSTEDNTDEYFSAINTNTVGDSSC-EDIETASNEENLRICEKM 859
QY 250 NTISAEVDSLLALLKESQAPIQGQIGOPTLPNNSLATPSPSLNPCTSHKHEEDGY 309
Db 860 N--EAMSSGDCVQND--GSKT--QISFSTSDPD---NFQESNDNTEFST 904
QY 310 GPDA-NRIIADESGFVMSHGDSNHYFFKKDLTEQIKAAKHLEEVKTSNGLDLSL 368
Db 905 KYKVRNSDLEDDDES-----LKKELTKAEV--VDKLDDEB-----ESEDY 941
QY 369 EODY-----PCNAK-----EMKLDKKIEBK---AGIMKQYGVTR 401
Db 942 EQDYADPEFGNDEGSNENIVKTKDITLIVPEPEKVKVHEEETLFEANVSSSYNVQN 1001
QY 402 ESI---VNVKEKNALITPSGDHSH-----ADPIDEKPVGIGHSHSNYELF 444
Db 1002 KMHVTVINQEAQA-NYEGEKYIYQNTDTEAHISIIERIDEN---AIG---NNKEI- 1053
QY 445 KPEEGVAKGKNRYTGEELTNVYN--LKNSTFNQNFILA-----NGKRVSF 492
Db 1054 -PERSCVKTHNEVLFERRATTENTKALENTNMHDQVQSACSDSDRQDQSTAEKNVGS 1112
QY 493 SPFPELEKLGIMNL--VKLIPDGVLEKVG--KVFG--EGVGNIANFELDQYPLPG 545
Db 1113 SAKHNDIRVSEISESVPLAPESDRSIFSPIRIVGAVKVGKVDV-----A 1164
QY 546 QPKFYIASKDPEVBYDCTFTVPTSLAYKMASQTFYFPFHAGTILRVNPOPAVPGTD 605
Db 1165 ESF---VKIDVMSDSDNDVIGD-----YNQDIFNSNSTDASVNMKSVSSKERDSD 1215
QY 606 ---ALVRVFEFHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLN 660
Db 1216 EDEAVILGGVTAEAHND--GNNSRVINI-----DPTNGAYEEDSEVFQVY-KDK 1264
QY 661 QSTYIIVEVPIL-----KENQDKPSI----- 682
Db 1265 ENLHSEPLVGELOEQHEFKDHSNEEEFTIYGDITSANIHSNAPDDIKRQOLLKN 1324
QY 683 -----LPQFRKKAQENS-KLDEKVEPKTSEK-----VEKEKLSGTNS 721
Db 1325 LSDLENYSQLIEDSRGNQESDEVNTSREDLTFEASVNEKYAGAEEDTFSELDIS 1384
QY 722 TSNSTLEVPVDPVQEK-----VAKFAESYGMKLE-----NVLNMDGTIELYIP 767
Db 1385 IQRPEHEEDLSDNNOERSTELNSPEEAELYELEIEGTETATAASSKKNDDERQGNIP 1444
QY 768 SGCEVIKKNAD---FTGEAPOGNGENKPNKGVSTGTVENOPTENKPADSLPEAPNEK 824
Db 1445 STDLPSDPPSDEEVTDSYPYNSENITAE-----KSAPTSPEVETIFSDTFNEVP 1495
QY 825 VK-----PENSTDNCKMPEGNVGDPMDDPA 851
Db 1496 MEINDEIPATLLEKHDKNTVSVLDDRSHELSDHVDNPHONSINKVNEGEPP----- 1550
QY 852 LEAPAVD-----PVQEKLEKFTASTGLDLSVIFNMGTIELRPLSGEVIKKNLS 902
Db 1551 --EHOAVDIPVKVEKEQOEEMPSK-----SVLEQKQPSMELINDKSS 1591
RESULT 7
BUD3_YEAST

ID BUD3_YEAST STANDARD; PRT: 1636 AA.
AC P25558; P25557; P87007;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bud site selection protein BUD3.
GN BUD3 OR YCL014W OR YCL14W/YCL13W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95247824; PubMed=7730410;
RA Chant J., Mischke M., Mitchell E., Herskowitz I., Pringle J.R.;
RT "Role of Bud3p in producing the axial budding pattern of yeast.";
RL J. Cell Biol. 129:767-778(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RA Staveva L.I.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Gromadka R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP REVISIONS.
RA Mewes H.-W.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CO-ASSEMBLES WITH BUD4 AT BUD SITES. BUD4 AND BUD3 MAY
CC COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
CC DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
CC ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U17580; AA86315.1; -
CC EMBL; X59720; CAA42346.2; -
CC PIR; S19341; S19341.
CC PIR; S19340; S19340.
CC SGD; S0000520; BUD3.
CC InterPro; IPR000219; RHOGEF.
CC SMART; SM00325; RHOGEF; 1.
KW Cell cycle.
SQ SEQUENCE 1636 AA; 184717 MW; 9E4E46BA5C3A3F69 CRC64;
Query Match 3.9%; Score 183; DB 1; Length 1636;
Best Local Similarity 19.0%; Pred. No. 0.29;
Matches 186; Conservative 146; Mismatches 316; Indels 332; Gaps 50;
QY 5 YDDEIQVAKLAGKYTDEGYIFD-----TSWIKKDSLSEAEARAAQAY 48
Db 690 HDDKHIEV-----TSNIVFTIINQIAIEIPICFSSLNSSMAKDLVCYENLNLE 741
QY 49 AKEKGLTPSTDHQDSGNTKGA-----EAIYNRVKAARKVPLDRLPYNTQYTYVEYKNSL 105
Db 742 HQLEEVKHPSTDEHRAVNSKLSGASDFDTHKKRSGYGTITTF-----SYTSLDKDS-- 794
QY 106 IPSYDHYHNKFEWDEGLYEAPKGYSLDLE---ATKYIYVEPR-----NLS--D 152
Db 795 --PSGDNSNVTK-----ETKEILFVPTKSKSKKPREIOKTKTNASKAE 837
QY 153 HVRKNKADO-----DSKPDDEKHEHVESEPTHPSEDEKENHAGLNPSAD 196
Db 838 HIEKKAPNKGKGFVGLKNVFGSKSKSPVORVFKISQRHPKSPVK----- 886

197 NLYKPTDTEETEEAEEDTDEAETPGTPTIRQANMETLTGLKSSLLGTDKNNNTISAEV 256
 887 ---KPTSEKK---SSPRVAVSPKIK-----KLSFSTKESQAKS-- 924
 257 DLSLALKESOPAPITQPGIOQPTLPNNSLATPS-----PSLIPNGTSHKHE 305
 925 -SLRAVEFKSDLLIGKPPDVNGAHQPOENTRISVVDRTKVSYNPSQPTVENTSNEKNV 983
 306 EGYGFDANRIAEDE-SGFVMSHGDSNHYFFKDL-TEEQI---KAQK-----H 351
 984 EP-----KADQSTKQDNISNFADEVSAVSPEKLDATDDQIIGKATNSSSVHGKNELPD 1039
 352 LBEVKTSHUNGDSLSHSDYPCNAKEMKDL-----KKEEKIAG-----IM 394
 1040 LAEVTA-NRVSTTSAGQDRTDQSEFLRADVENLSDDDHQRNESVFNDLFGDFIP 1098
 395 KOYGKRESIVVNEKNAIYPSGDHHDPIDEHKPVGIGHSHSNYELFPEGVAKKE 454
 1099 KHYRNQENI--NSSN--LPPEG-----KVPQEK--GVSNENTNISLKTNEADASTLTQ 1146
 455 GNKVTGELTNVNLNNSFNQ-----NFTLANGQKRVFSFPPE-----497
 1147 KLSPOASKVLTNSNELKDT--NNEGDKADIKLGGDDYSDEKETAKEIT---KPKNFVSGI 1201
 498 LEKLGINKMLKLTIPDGKVLKESGKRVGEGVGNIAFELDQPYLPQGTQKTYTASDY 557
 1202 TERKEIPTIPLAPPASKI-----NFQSPSYI---ELFGQMRVLDK 1242
 558 PEVSVDGFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAPVPGTDLALVYVDFEHN 617
 1243 HDHYNMK-----RLASQVSL-----SEGLKVN-----TEEDAAIINKSQDD 1279
 618 AYLENNYKVEI-----KLPIPKLNGQTRTAGNKIPTFMANAYLDNQSYIVVEPTLE 672
 1280 AKARMTQISEIYETEMOQPI-----TYLPKALHDSG-----IE 1315
 673 KENOTDKPSILPQKRNKAQENSKLDEKVEPKTSEKKEKLSGTGNSNSTLEPYPT 732
 1316 KSD-----DREFEELKELKSKGTGNEGVNNFNSIPKLEKPPA 1359
 733 -----VDPVQEKVAFESYGMKLEN-----VLFND-----GTIELVLPSEGVIK 773
 1360 FKVINTSPVRIIGTFEDT--RKYENGSPDISFTYTHNNDPEPKRLMELKFPQDEIP 1417
 774 KNAADFTGAPOGNGENPSENGKYSTGTVENQPTENKPADSLPEAPNEKPKVKN--ST 831
 1418 DRRF-YT-----PAE-----EPTAEPTVEELPNTPRSNVTTNNKST 1454
 832 DNGMLNPEGNVSDP--MID 849
 1455 DDKL--SSGNIDQKTELD 1472
 RESULT 8
 ID HWML MYCGE STANDARD; PRT: 1139 AA.
 AC Q4913; Q49365;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytochrome high molecular weight protein 1 (Cytochrome accessory
 protein 1).
 GN HWML OR MG312.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_Taxid=2097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,

Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 Tomb J.-P., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
 Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
 "The minimal gene complement of Mycoplasma genitalium".
 Science 270:1397-403(1995).
 [2]
 SEQUENCE OF 721-847 FROM N.A.
 STRAIN-ATCC 33530 / G-37;
 MEDLINE=94075230; PubMed=8253680;
 Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III,
 "A survey of the Mycoplasma genitalium genome by using random
 sequencing".
 J. Bacteriol. 175:7918-7930(1993).
 CC -FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
 CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
 CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
 CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
 CC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
 CC SIMILARITY).
 CC -SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
 CC MEMBRANE (BY SIMILARITY).
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL; U39712; AAC1534.1; -;
 CC EMBL; U02261; AAD12527.1; -;
 CC TIGR; MG312; -;
 CC Cytochrome; Structural protein; Complete proteome.
 CC Cytadherence; Structural protein; Complete proteome.
 CC
 CC Query Match 3.8%; Score 181.5; DB 1; Length 1139;
 CC Best Local Similarity 19.7%; Pred. No. 0.21;
 CC Matches 161; Conservative 118; Mismatches 342; Indels 197; Gaps 39;
 QY 146 EPRNASDHVRKADQSKPDKEHDEVSEPTPESEKHNAGLAPSDNLKPTDT 205
 DB 18 EPNIFGDLGKSTVEEDPNKIVAYD-----ADNGNYIYAFNKTGVYDPTGDT 68
 QY 206 EETEEAEEDTDEAETPGTPTIRQANMETLTGLKSSLLGTDKNNNTISA-----E 255
 DB 69 EYDISQLEDN-----GNPFTDEKQENDYLK---YVGNPDYGSIDENGSEWWSGVFE 119
 QY 256 VDSLALALKESOPAPIQ-----GPQIGQPTLPNNSLATPSPLIPNGTSHKHEEDG 308
 DB 120 NDQWIS-TKESQPTDENYGFSDLPPEVKQPEVEDNYGFDN-DLP--PEVKQPESEVDN 175
 QY 309 YGFDANRI---IAEDSGFVMSHGDSNHYFFKDLTEEQIAAKQKHEEVTSHNGDLSL 365
 DB 176 YGFD-NLPPPEVKQPE--VVDQSSDDYFAKQP-TDENYGFNDLPPPEVKQPEVVDQ 231
 QY 366 SS--HEQDYPGNKEMKDLKKEIKIAGINKQYGVKRESIVVNEKNAIYPSGDHHA 423
 DB 232 SSDHFAKQPESTDSYFSDSLDLPQ-----TLQPSLDHVVQ 269
 QY 424 DPIDSH---KPVGIGHSHSNYELFPEGVAKKGNKYVTGEELTNVNLNKNSTFNQ 480
 DB 270 YNFDHHEELKPV--AEQNNYQV-----GFDQVQAN-LDNNEEIQPTAEKKVYTFDESQ 321
 QY 481 FTLANGQKRVFSFPPELEKLGINKMLKLT-PPGKVLKESGKRVGEGVGNIAFELD 539
 DB 322 AQVVD-----SYQLPIDTDQDDQTTFFSSFTQPTVEQFDQVNSEV-----362
 QY 540 QYLPQGTQFTKTIASKDYPEVSDGTFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQ 599
 DB 363 -----NDQFKPEITKEPVLESFNFQDVVTS---DLNSENLSYENKKA-----405
 QY 600 VPKGTDALVRV-----DEFHGNAYLE---NNYKVGKIKLPIPKLNGQTRTAGN 646

Db 406 --TNDLSNFIQLNSSETASDDVHYESKSEPHDKFGS-----DLQSNNSNLSLE 457
 Qy 647 KIPVTFMA-----NAYLDNQSYI--VEVPILLEKEN---QTKRPSILPOFKRKAQENSKL 697
 Db 458 SEPVKFNSETAPDAHFSEQSEFVQOYDIYQNEELKPTLDQPSDDYFAKQPTDNTGFG 517
 Qy 698 DEK-----VEEPTSEKVEKELSTGNSTSLTEEVPTDP-----VQ--- 737
 Db 518 DNDLPEVKQESVVDQSSDDHFAKPESTDSYFSDLDLPQ--PTLDQPSLDDHVQNF 576
 Qy 738 -----ERKVAFAESYGMKLENVLMQMDGTIELYLPSEGVIRKKNMADFTGEAPQNGENK 791
 Db 577 DHHEELKPAVEENQYQVGFQVQANLNDNEEQ--PTAE--KEVTTDFESKQAQ----- 627
 Qy 792 PSNGKVTGTVENOPTENKPADSLPEAPNEKPKVPENSTDMKLNPNVGVSGDDPMDPA 851
 Db 628 VYDSQLPIDT--DOODQTTSSSFETQPTVEQFQVNSEVNDQPKPE--ITKEPVLESS 683
 Qy 852 LEEAPAVDPQVEKLEKFTASYGLGLDSVFNMDGTIEL 889
 Db 684 FNKQDVVE-----TGNNTNNLQKFDIQSDNKITI 712

RESULT 9
 CNA_STAAU STANDARD; PRT; 1183 AA.
 AC Q53654;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen adhesin precursor.
 GN CNA.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FDA 574;
 RX MEDLINE=92165839; PubMed=1311320;
 RA Patti J.M., Jonsson H., Guss B., Switalaki L.M., Wiberg K.,
 RA Lindberg M., Hoecek M.;
 RA "Molecular characterization and expression of a gene encoding a
 RT Staphylococcus aureus collagen adhesin."
 RL J. Biol. Chem. 267:4766-4772(1992).
 RN [2]
 RP ERRATUM.
 RA Patti J.M., Jonsson H., Guss B., Switalaki L.M., Wiberg K.,
 RA Lindberg M., Hoecek M.;
 RL J. Biol. Chem. 269:11672-11672(1994).
 RN [3]
 RP COLLAGEN-BINDING DOMAIN.
 RC STRAIN=FDA 574;
 RX MEDLINE=94032261; PubMed=8218209;
 RA Patti J.M., Boles J.O., Hoecek M.;
 RA "Identification and biochemical characterization of the ligand
 RT binding domain of the collagen adhesin from Staphylococcus aureus."
 RL Biochemistry 32:11428-11435(1993).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
 RX MEDLINE=97475225; PubMed=9334749;
 RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
 RA Moore D., Jin L., Schneider A., Delucas L.J., Hoecek M.,
 RA Narayana S.V.L.;
 RA "Structure of the collagen-binding domain from a Staphylococcus
 RT aureus adhesin."
 RL Nat. Struct. Biol. 4:833-838(1997).
 CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
 CC COLLAGEN-CONTAINING SUBSTRATA.
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (Potential).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; M81736; AAA20874.1; --
 DR PDB; 1AMX; 24-JUN-98.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR TIGRFAMs; TIGR01167; LPTXG_anchor; 1.
 DR PROSITE; PS50847; GRAM_POS_ANCHORING; FALSE_NEG.
 KW Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
 FT SIGNAL 1 29
 FT CHAIN 30 1154
 FT PROPEP 1155 1183
 FT DOMAIN 151 318
 FT DOMAIN 533 1093
 FT DOMAIN 1093 1157
 FT REPEAT 533 719
 FT REPEAT 720 906
 FT REPEAT 907 1093
 FT SITE 1151 1155
 FT MOD_RES 1154 1154
 SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E57D76 CRC64;
 Query Match 3.8%; Score 180.5; DB 1; Length 1183;
 Best Local Similarity 21.2%; Pred. No. 0.24; Indels 261; Gaps 46;
 Matches 198; Conservative 122; Mismatches 351;
 Qy 46 QAYAKEK--GLTPPSTDHQDQNTAKGAEIYNVKAARKVPLDRMPY--NLYQTVVEKN 102
 Db 300 QEHGEEVNGKSPNHTVHNANAGIEG--TVKGLKVLKQDKTKAPIANVFKLSKKD 357
 Qy 103 GSL-----IIPSYDHYHNKFEWDEGLY-----EAPGYSELDLLATVKTVEPR 148
 Db 358 GSVVKNQKQKEIITDANGIANIKA--LPSGDYILKIEAPRYTFD----- 402
 Qy 149 NASDHVRKNKADQSKPDDEK-----EHDEVSPETHPESEK--ENHAGLNPSA-DNL 198
 Db 403 -----KDEYPTMKDQDNGYTTIENAKAIEKTDVSAQRWEGTQVKPYIYFKL 455
 Qy 199 YKPSTDETEEEAEDTT--DEAEI-----PGTSPISRONAMETLTGLKSLSLGTLKDNNTI 252
 Db 456 YK-----QDDNQNTPTVDKAEIKLEDGTTKVTW-----SNLPENDKNGRAI 497
 Qy 253 SAEVDSLLALKESOPAIQGIQPTLPNNSLATPSPSLPI--NPGTSHHEHEDEGTF 311
 Db 498 KYLVKEVNAQGEDTTPGY-----TKENGLVVTNTEKPIETTSISGEKVVDDKQ 549
 Qy 312 DANRIIADESGFVMSHSDSNHYFPFKDLTEQIKAAQKHLEEVKTHSHGLDLSHSD 371
 Db 550 DGR--PERKSVNLLANG-----EKVT-----LDVTSETNWK 580
 Qy 372 YPGNAKEMKDLDDKIE--EKIAGIMKQYGVK-----RESIVYNNK--EKNAIIPSGDH 420
 Db 581 Y-----EPRDLPKYDEGKIEVTVEDHVKYDVTDDINGTTITNKYTPGTSTATVKWDD 635
 Qy 421 HHADPIDEHKPVGIGHSHSNYELFPBEGVA-----KKEGKV 458
 Db 636 NNNQ--DGRKPTTEI-----KVELY--QDGKATGKATILNESNNWTHWTGLDEKAGQOV 686
 Qy 459 -YTGELTNVNLKSTFNNOFTLANQKRVSFPPPELEKLGILNVLKITPDGKV 517
 Db 687 KTVSELTKV-----KGYTHVDNNDH--GNLIVTNKTPETTSISGEKVVDDKQDGR 740
 Qy 518 LEKVSQKVFEG-----VGNIANFELDQPLP-----QTEFYTIAS---KDYPEVSYD 563
 Db 741 PEKSVNLLADGEKVKTLDTSETNWKYEFKDLPKYDEGKIEVTVEDHVADY--TTDIN 799
 Qy 564 GTFTVPTSLAYKMASQTIFFYPFHAGDTYLRV-----NPQFAVPKGTDLALVRFDE---- 613
 Db 800 GT-----TINKYTPGETSATVTKWDDNNNDKQDKRPTTEIKVELYQDGRAT 845

	EMBL; X02919; CAA26676.1;	-
DR DR	PYR; A24594; A24594.	
DR	InferPro; IPR000561; EGF-like.	
DR	Ffam; PF00008; EGF; 1.	
KW KW	Malaria; Merzoite; Polyprotein; Repeat; Signal; Glycoprotein;	
KM	Transmembrane; GPI-anchor.	
FT SIGNAL	1	19 POTENTIAL.
FT CHAIN	20	MEROZOITE SURFACE PROTEIN 1..
FT CARBOHYD	116	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	268	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	764	N-LINKED (GLCNAC. .) (POTENTIAL).

FT. CARBOR

RESULT 11

SC16_YEAST STANDARD; PRT; 2195 AA.

AC P48415, 002822; Rel. 33, Created

DT 01-FEB-1996 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE Multidomain vesicle coat protein.

GN SEC16 OR YPL085W OR LPLW.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96017704; PubMed=7593161;

RA Espenshade P., Gimeno R.E., Holzner E., Teung P., Kaiser C.A.;

RT "Yeast Sec16 gene encodes a multidomain vesicle coat protein that

RL interacts with Sec23p.";

RL J. Cell Biol. 131:311-324(1995).

RN [2]

RP SEQUENCE FROM N.A.

RX STRAIN=S288C / AB972;

RA MEDLINE=97313271; PubMed=9169875;

RA Bussey H., Storms R.K., Ahmed A., Albertmann K., Allen E., Ansoorge W.,

RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,

RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,

RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,

RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,

RA Duncan M., Floeth M., Fortin M., Friesen J.D., Fritz C., Goffeau A.,

RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,

RA Hunkeler-Smit S., Hyman R., Johnston M., Kalman S., Kleine K.,

RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,

RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,

RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,

RA Petek F.X., Pohl T.M., Purnelle D., Schaefer M., Scharfe M.,

RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,

RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,

RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,

RA Zhong W.W., Zollner A., Vo D.H., Hani J.;

RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";

RL Nature 387:103-105(1997).

CC -!- FUNCTION: INVOLVED IN THE BUDDING OF TRANSPORT VESICLE FROM THE

CC ENDOPLASMIC RETICULUM. THE C-TERMINAL INTERACTS WITH SEC23 AND

CC WITH THE CYTOSOLIC DOMAIN OF SED4. COULD THEREFORE BE A

CC CONSTITUENT OF COPII VESICLE COAT. N-TERMINAL OVEREXPRESSION

CC CAUSES A LETHAL SECRETION DEFECT.

CC -!- SUBCELLULAR LOCATION: ON THE ENDOPLASMIC RETICULUM AND ON VESICLES

CC WHICH BUD FROM IT.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC

DR EMBL; U23819; AAC49088.1; -

DR EMBL; U41849; AAC68254.1; -

DR SGD; S0006006; SEC16.

KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum.

FT DOMAIN 1997 2094

FT MUTAGEN 1059 1059

FT L->S: IN SEC16-4; TS ACCUMULATION OF ER

FT MEMBRANES.

FT L->P: IN SEC16-3; TS ACCUMULATION OF ER

FT MEMBRANES.

FT L->P: IN SEC16-2; TS ACCUMULATION OF ER

FT MEMBRANES.

FT W->R: IN SEC16-1; TS ACCUMULATION OF ER

FT MEMBRANES.

FT MUTAGEN 1231 1231

FT MUTAGEN 1084 1084

FT MUTAGEN 1089 1089

FT MUTAGEN 1089 1089

FT MUTAGEN 1231 1231

FT MUTAGEN 1084 1084

FT CONFLICT 522 522

FT CONFLICT 560 560

FT MISSING (IN REF. 1).

FT I -> F (IN REF. 1).

SQ SEQUENCE 2195 AA; 241694 MW; 757B7A7231BEE6F0 CRC64;

Query Match 3.7%; Score 176.5; DB 1; Length 2195;

Best Local Similarity 21.8%; Pred. No. 0.86;

Matches 203; Conservative 113; Mismatches 371; Indels 243; Gaps 49;

QY 28 DTSWIKDLSSEARAAAYAKAKGLTPPSTHDSQNTGTEAK-----GAEATYNRVK 80

DB 1393 NTGWLAKPKLSVWGQLDKSPNKYIG--GDDIALNKKNDKKKVFDTGGTSGSSANSVD 1450

QY 81 AAKVPLDRMPYNLYTVE--VKNGSLI-----IPSYDHYHNKFEWDEGLYEA--PKG 131

DB 1451 LTQTF-----TPFAQVTSQSYDVTALLNAHNVPVSHVLS--KPSNVSKGLVEANLPTT 1505

QY 132 YSLDLLATVKYIVPRNASDVRKADQSKPDDEKDEHDEVSEPTHPDESDEKENHAGL 191

DB 1506 HRIGDSL-----QGS-----PORLHNTQFAAA 1527

QY 132 NPSADNLYKPTD--TEETEEAEADTTDEAEIPGPFSIRQN---AMETLTGLKSSLL--- 243

DB 1528 EPQMASLRVRTDQHTNEKALKSQILEKKSTAYTPQGCQNHSPVMEKSNVPSLEADF 1587

QY 244 ----LGTQDNNTISA-----EVDLSLALLKESQAPIQGPQIGQPTLPNNS----LATPS 290

DB 1588 PAPPKLGTVPSNYVSSPDLVRRESIISTGSEFLPP-----PKIGVPTKANSQGSLSMYS 1643

QY 291 -PSLPINPOTS--HEKHEEDGYGFANRIIADES--GFVNSHGDSNHYFFKKDLTTEEQI 345

DB 1644 VVALIDPVPVQVHETGYNDFOKHSQKSMPEDESHTSHDNSNAQONTLKDSADVTDET 1703

QY 346 KAAQKHELEVTSNGLDLSLSSHEQDYPGNAKEMDKDKKIEKTAGIMKQYGVKRETV 405

DB 1704 DIEGPGENDVK--NLLPMPNHPQTSTVNPVPIQ-----TISDDIOPIL-QTNVEVRGT 1753

QY 406 VNKENALYIPSGDHHADPIDEHPGVGHSHSNHYELFKPEGVAKREGKNKYVGEELT 465

DB 1754 ASKEMNSL--PSIENERS---SEEQPENISKSASS--AYLPSTGGLSLE-NRPLTQDE-- 1803

QY 466 NVVNLKKNSTPNQNTLANGOKRVSFSPPELEKKLGINMLVKLITPDGKLYEKVSGV 525

DB 1804 NSIETVQSTY-----LPAG-----SISMEAKPISQVDVPRVNNKA 1841

QY 526 FGEQGVGNIA-----NFELDQPYLPQGTQTKYTIASKDYPEVSYDGTFTVPTSLA 573

DB 1842 SKLVEQHMAPPKPKSTDAIKMY---SPVTP-----QSTAASADGDE---STILKTSPAI 1890

QY 574 YKMASQ-----TIYFPF--HAGDTYLRVNPQFAYPKGTDALVRVDFEFGHAYLENNYK 625

DB 1891 YARTHQAHASNPQYFPLVNOANET-----ASFELSESTSQ-----QSGNHNVAENRFS 1940

QY 626 VGEIKLPKPLN-----QGTRTAGNK-----IPVTFMANAYL-----DNQSTY 664

DB 1941 -----PIKAEVVEKDTFQPIRKASTNQYAFPLESDADKYNQVIEDESDNNKSTD 1994

QY 665 IYEVPILEKENTDKPSILPQPKRNAQNSKLDEK-----VEEPTSEKVEKEKLSSE 717

DB 1995 EAKNRKEEKKNVMMKTKP-----SNKDIDDKSNWFGWLKKDKTKGKVKYAKLGH 2046

QY 718 TGNSTNSNSTLEEVPVDPQVQEVAKFAESYGMKLENLNFMD-----GTIELYLP- 767

DB 2047 KNTLYYDEKLKRWNVNKDTEEEKQKIIESSAPPPPIVKKDKGGPKTRPSGPPINNLSPP 2106

QY 768 --SGEVIKKNMADFTGE-----APOG--NGENKPSNGKVS--TGTVENOPTENKPAD 814

DB 2107 VVATSVIPNN--PIYGEPLPIKTSPTGPNPNPNPSPSPISRISGVNLTSEKKNGLDD 2164

QY 815 SLPEAPNEKP--VKPENSTDNGLNPEGVN 842

DB 2165 LLSLAGGPKPASTRRKKTKARGVYVNVNDNI 2194.

RESULT 12

PGCV_RAT

[illegible]

1423 TVD--LTENRDIENVNSTMVDLPQTMEPAKLWSKPE-----VNFKEQIEGSE 1467

RESULT 13

ANK2_HUMAN	STANDARD;	PRT; 3924 AA.
ID	ANK2_HUMAN	
AC	Q01484; Q01485;	
DT	01-APR-1993 (Rel. 25, Created)	
DT	01-OCT-1996 (Rel. 34, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).	
GN	ANK2.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
PC	[1]	
PP	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).	
RC	TISSUE=Brain stem;	
RC	MEDLINE=91302466; PubMed=1830053;	
RA	Otto E., Kunimoto M., McLaughlin T., Bennett V.;	
RA	"Isolation and characterization of cDNAs encoding human brain	
RT	ankyrins reveal a family of alternatively spliced genes.";	
RL	J. Cell Biol. 114:241-253(1991).	
RL	[2]	
RN	SEQUENCE OF 463-495 FROM N.A. (ISOFORM 1).	
RP	Carpenter S.;	
RP	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.	
RL	[3]	
RN	SEQUENCE FROM N.A. (ISOFORM 1).	
RC	TISSUE=Brain stem;	
RC	MEDLINE=94075409; PubMed=8253844;	
RA	Chan W., Kordeli E., Bennett V.;	
RA	"440-kD ankyrin: structure of the major developmentally regulated	
RT	domain and selective localization in unmyelinated axons.";	
RL	J. Cell Biol. 123:1463-1473(1993).	
RL	[4]	
RN	SEQUENCE OF 463-495 FROM N.A.	
RP	MEDLINE=92009921; PubMed=1833308;	
RP	Lee W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,	
RA	Lux S.E., Ward D.C., Forget B.G.;	
RT	"Isolation and chromosomal localization of a novel nonerythroid	
RL	ankyrin gene.";	
RL	Genomics 10:858-866(1991).	
CC	-1- FUNCTION: Attach integral membrane proteins to cytoskeletal	
CC	elements. Also bind to cytoskeletal proteins.	
CC	-1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE	
CC	PRODUCED BY ALTERNATIVE SPLICING.	
CC	-1- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL	
CC	CELLS THROUGHOUT THE BRAIN.	
CC	-1- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES	
CC	AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE	
CC	AND FUNCTION (POTENTIAL).	
CC	-1- SIMILARITY: CONTAINS 23 ANK REPEATS.	
CC	-1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.	
CC	-----	
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; X56957; CAA0278.1; -	
DR	EMBL; X56958; CAA40279.2; -	
DR	EMBL; Z26634; CAA62644.1; -	
DR	EMBL; M37123; AAA62828.1; -	
DR	PIR; S14533; S14533	
DR	PIR; A39643; A39643	
DR	PIR; B39643; B39643	
DR	PIR; S14569; S14569	
DR	HSSP; P42771; IDC2.	

QY 89 RMPYNLOTVTEV-----KNGSLIIPSYDHYHNIKFEW-----FDEGLYEAP----- 129
Db 1521 RVREDLEKNEILRSCTCTDESSVQSSRGLVEEENWIVDSIEEERARQKAPLEITE 1580
QY 130 -----KGYSDLELLATVYKYPERN-----ASHVRK----- 156
Db 1581 YPCVEVRIDKEIKGKEDKSTGLVNTLTDLCVLPKPEQLQTVQDKAGKCEALAVGR 1640
QY 157 --NKADDSKPE-----DREHDE---VSEPTHPESDE---KENHAGLNPSADNLYKPESTD 205
Db 1641 SSEKEGKDIPDETQSTQKQKSLGKIKKPVRRKLLKQKKEGLQASAKAELEKGGSS 1700
QY 206 EETEEEAEDTDEAEIPGTPSIQONAMETITGLK--SLLLGTKDNNTTISAEVDSLLALLK 264
Db 1701 EES-----LGEDPGL-----APEPLTVKATSPLEETPIGSIKDKVKALAKRVE 1745
QY 265 EQ-----PAPIOGPO-----IGQPLPNNSLATPSP----- 292
Db 1746 DEQGRSKLPVRKGVKEDVPKTTTHRPHPAASPGLSERHAPGSPSPKTERHSTLSSAK 1805
QY 293 ----LPINPGTSHKHEEDGVGFANRIIAEDESGFVNSHG-----DSNHYFFKNDLT 341
Db 1806 TERPPVSPSPSKTEKHPSPVSPSAKTERHSPASSSSKTEKHPSPSPSTKTERHSPVSTKT 1865
QY 342 EEQ--IKAAQKHEEVKTSNGLDLSLSSHODYPGNKAKEMKDLDDKKEEIKIAGIMQYGV 399
Db 1866 ERHPPVSPSGTKDRPPVSPSG--RTEKHPVSPSGT-----EKRLPVSPSGRTDKH-- 1915
QY 400 KRESIVYKNEKNALIPSGDHHADPID-EHKPGVIGCHSHSNYELFKP----- 446
Db 1916 QPVSTAGTEKHLVPSPSGTEKQPPVSPSTKTRIEETMSVRELMAKAFQSGQDPSSKHT 1975
QY 447 ---EEGVAK---KEGNKVTGELTNVYNLLKNSTNNONFTLANGOKRVSFSPFPELE 499
Db 1976 GLFEHSAKQKQKQKGVKVEKEKGP--LTQREAOQNTQIKRGOR---LPTVGTAE 2030
QY 500 KKLGINMLVLIITPDGKLVKSVKGVGEGVGNANTANLDPPLPGQTFKTIASKDYPE 559
Db 2031 SKRGVR-----VSSIGVKKEDAAG-----GKEVLSHKIPEPV-----QSVPE 2088
QY 560 VSYDGTFTVPTSLAYKMASQITFYPPFHAGDTLYRVNPOFVAPKGTDALVRVDFEFGNAY 619
Db 2089 EESHRESVPRK---KMADE-----QGDMDLQISPRDKTSTDFSEVIK-----QE 2110
QY 620 LENNYKVGKIKLIP-----KLNGTTRTAGNKIPVTFMANAY-----LDNOST 663
Db 2111 LEDNOKYQQFRLSEETAKAQLHLDQVLTSPNTTFLDYMKDEFPLALSLQSGALDSS- 2169
QY 664 YIVEVPILKENQTDKP--SILPQKRNKAQENSKLDEKVEEPT-----SEKVEKEK 714
Db 2170 -----ESLKNEGVAGSPGSLMEGTPOISSSESYKHEGLAETPETSLSFSPPKSEQ 2224
QY 715 LSETGNSTNSLTLEV-----PTVDPV---QEKVAKFAESYGMKLENVLFNMDGTIEL 764
Db 2225 TGETESTKTTTETIRSEKEHPTTKDITGGSEERGATVTEDESETSES--FQKATLIG- 2281
QY 765 YLPSGEVTKNNMADTG-----EAPGNGENKTPSNGKRVSTG--TVENOPTENKPADS 815
Db 2282 --SPKDTSPKRDQDCTGSCSVALAKETPTGLTEEAACDEGQRTFGSSAHKTQTDSEAQES 2339
QY 816 -----LPEAPNEKVPKPNSTONGMLNPNNGVSDPMLDPA 852
Db 2340 TATSDTKALPULPEA---SVKTDGTGES---KPOGVINSPOGLELAL 2380

RESULT 14
L778_ARATH
ID LT78_ARATH STANDARD; PRT: 710 AA.
AC Q06738; Q04981;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

QY 89 RMPYNLOTVTEV-----KNGSLIIPSYDHYHNIKFEW-----FDEGLYEAP----- 129
Db 1521 RVREDLEKNEILRSCTCTDESSVQSSRGLVEEENWIVDSIEEERARQKAPLEITE 1580
QY 130 -----KGYSDLELLATVYKYPERN-----ASHVRK----- 156
Db 1581 YPCVEVRIDKEIKGKEDKSTGLVNTLTDLCVLPKPEQLQTVQDKAGKCEALAVGR 1640
QY 157 --NKADDSKPE-----DREHDE---VSEPTHPESDE---KENHAGLNPSADNLYKPESTD 205
Db 1641 SSEKEGKDIPDETQSTQKQKSLGKIKKPVRRKLLKQKKEGLQASAKAELEKGGSS 1700
QY 206 EETEEEAEDTDEAEIPGTPSIQONAMETITGLK--SLLLGTKDNNTTISAEVDSLLALLK 264
Db 1701 EES-----LGEDPGL-----APEPLTVKATSPLEETPIGSIKDKVKALAKRVE 1745
QY 265 EQ-----PAPIOGPO-----IGQPLPNNSLATPSP----- 292
Db 1746 DEQGRSKLPVRKGVKEDVPKTTTHRPHPAASPGLSERHAPGSPSPKTERHSTLSSAK 1805
QY 293 ----LPINPGTSHKHEEDGVGFANRIIAEDESGFVNSHG-----DSNHYFFKNDLT 341
Db 1806 TERPPVSPSPSKTEKHPSPVSPSAKTERHSPASSSSKTEKHPSPSPSTKTERHSPVSTKT 1865
QY 342 EEQ--IKAAQKHEEVKTSNGLDLSLSSHODYPGNKAKEMKDLDDKKEEIKIAGIMQYGV 399
Db 1866 ERHPPVSPSGTKDRPPVSPSG--RTEKHPVSPSGT-----EKRLPVSPSGRTDKH-- 1915
QY 400 KRESIVYKNEKNALIPSGDHHADPID-EHKPGVIGCHSHSNYELFKP----- 446
Db 1916 QPVSTAGTEKHLVPSPSGTEKQPPVSPSTKTRIEETMSVRELMAKAFQSGQDPSSKHT 1975
QY 447 ---EEGVAK---KEGNKVTGELTNVYNLLKNSTNNONFTLANGOKRVSFSPFPELE 499
Db 1976 GLFEHSAKQKQKQKGVKVEKEKGP--LTQREAOQNTQIKRGOR---LPTVGTAE 2030
QY 500 KKLGINMLVLIITPDGKLVKSVKGVGEGVGNANTANLDPPLPGQTFKTIASKDYPE 559
Db 2031 SKRGVR-----VSSIGVKKEDAAG-----GKEVLSHKIPEPV-----QSVPE 2088
QY 560 VSYDGTFTVPTSLAYKMASQITFYPPFHAGDTLYRVNPOFVAPKGTDALVRVDFEFGNAY 619
Db 2089 EESHRESVPRK---KMADE-----QGDMDLQISPRDKTSTDFSEVIK-----QE 2110
QY 620 LENNYKVGKIKLIP-----KLNGTTRTAGNKIPVTFMANAY-----LDNOST 663
Db 2111 LEDNOKYQQFRLSEETAKAQLHLDQVLTSPNTTFLDYMKDEFPLALSLQSGALDSS- 2169
QY 664 YIVEVPILKENQTDKP--SILPQKRNKAQENSKLDEKVEEPT-----SEKVEKEK 714
Db 2170 -----ESLKNEGVAGSPGSLMEGTPOISSSESYKHEGLAETPETSLSFSPPKSEQ 2224
QY 715 LSETGNSTNSLTLEV-----PTVDPV---QEKVAKFAESYGMKLENVLFNMDGTIEL 764
Db 2225 TGETESTKTTTETIRSEKEHPTTKDITGGSEERGATVTEDESETSES--FQKATLIG- 2281
QY 765 YLPSGEVTKNNMADTG-----EAPGNGENKTPSNGKRVSTG--TVENOPTENKPADS 815
Db 2282 --SPKDTSPKRDQDCTGSCSVALAKETPTGLTEEAACDEGQRTFGSSAHKTQTDSEAQES 2339
QY 816 -----LPEAPNEKVPKPNSTONGMLNPNNGVSDPMLDPA 852
Db 2340 TATSDTKALPULPEA---SVKTDGTGES---KPOGVINSPOGLELAL 2380

Query Match
Best Local Similarity 19.5%; Pred. No. 2.4;
Matches 197; Conservative 125; Mismatches 400; Indels 286; Gaps 47;
QY 35 DSLSEAEERAAQYAKKGLTPSTBQDNGTEAGAEIYNRVRAAKKVP-----LD 88
Db 1469 DLLSEVSEMKQDLIKWAILT-----TVDSKA-----GSIKVKELVRAAEPEEPPEIVE 1520

Low-temperature-induced 78 kDa protein (Desiccation-responsive protein 29A).
 LTI78 OR LTI140 OR RD29A OR COR78 OR AT5G52310 OR K24W7.4.

Arabidopsis thaliana (Mouse-ear cress).
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 eurosids II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

SEQUENCE FROM N.A.

STRAIN=cv. Columbia; TISSUE=Leaf;

RC MEDLINE=93192524; PubMed=8448363;

RA Nordin K., Vahala T., Palva E.T.;

"Differential expression of two related, low-temperature-induced

genes in Arabidopsis thaliana (L.) Heynh.;"

PLant Mol. Biol. 21:641-653(1993).

[2]

SEQUENCE FROM N.A.

STRAIN=cv. Columbia;

RC MEDLINE=94143472; PubMed=8310052;

RA Yamauchi-Shinozaki K., Shinozaki K.;

"Arabidopsis DNA encoding two desiccation-responsive rd29 genes.;"

PLant Physiol. 101:1119-1120(1993).

[3]

SEQUENCE FROM N.A.

McLarny B.K., Gilmour S.J., Thomashow M.F.;

Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.

[4]

SEQUENCE FROM N.A.

STRAIN=cv. Columbia;

RC MEDLINE=30181125; PubMed=10718197;

RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,

Tabata S.;

"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence

features of the regions of 3,076,755 bp covered by sixty p1 and TAC

clones.;"

PLant Mol. Biol. 16:1061-1071(1991).

[5]

SEQUENCE OF 357-710 FROM N.A.

STRAIN=cv. Columbia; TISSUE=Leaf;

RC MEDLINE=91322497; PubMed=1830821;

RA Nordin K., Heino P., Palva E.T.;

"Separate signal pathways regulate the expression of a low-

temperature-induced gene in Arabidopsis thaliana (L.) Heynh.;"

PLant Mol. Biol. 16:1061-1071(1991).

CC -1- TISSUE SPECIFICITY: ACCUMULATES RAPIDLY IN LEAVES AND ROOTS.

CC -1- INDUCTION: BY WATER STRESS OR ABSCISIC ACID (ABA), AND MOSTLY BY

LOW TEMPERATURE.

CC -1- SIMILARITY: BELONGS TO THE LTI78/LTI65 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb.ch/announce/>
 or send an email to license@isb.ch).

CC EMBL; X67671; CAA47903.1; .

DR EMBL; D13044; BAA02376.1; .

DR EMBL; L22567; AAA32775.1; .

DR EMBL; AB019226; BAB10528.1; .

DR EMBL; X57600; CAA40826.1; .

DR PIR; S16273; S16273.

KW Multigene family; Repeat.

FT DOMAIN 63 67

FT 2 X 14 AA REPEATS OF P-[MV]-G-F-G-[DS]-

FT E-S-G-A-E-L-E-K.

FT 1-1.

FT REPEAT 303 316

FT 1-2.

FT REPEAT 357 370

FT 3 X 15 AA REPEATS OF [DN]-[FS]-P-[STV]-R-

FT S-H-[DE]-[FL]-D-[LM]-K-[NT]-E-[ST].

FT 2-1.

FT REPEAT 317 331

FT REPEAT 336 350 2-2.
 FT REPEAT 398 412 2-3.
 FT DOMAIN 510 600 5 X 5 AA REPEATS OF [FV]-[ADT]-[EST]-
 [KW]-L.
 FT REPEAT 510 514 3-1.
 FT REPEAT 532 536 3-2.
 FT REPEAT 550 554 3-3.
 FT REPEAT 579 583 3-4.
 FT REPEAT 596 600 3-5.
 FT DOMAIN 638 641 POLY-GLY.
 FT DOMAIN 648 696 2 X 23 AA REPEATS.
 FT REPEAT 648 670 4-1.
 FT REPEAT 674 696 4-2.
 FT CONFLICT 216 216 S -> P (IN REF. 2).
 FT CONFLICT 491 491 E -> V (IN REF. 2).
 FT CONFLICT 514 514 L -> H (IN REF. 5).
 SQ SEQUENCE 710 AA; 77856 MW; 9C6C8ACAE6BDF334 CRC64;

Query Match 3.7%; Score 173.5; DB 1; Length 710;

Best Local Similarity 19.9%; Pred. No. 0.26;

Matches 153; Conservative 82; Mismatches 271; Indels 263; Gaps 32;

QY 284 PINPOTSEKHEEDYGFDPANRIIAEDSGFVMSHGDSNHYFFKKDLTEQIKAAOKHLE 353
 DB 8 PLN---THQHPPE-----VEHENGATKMKR-----VKRAKFK 41
 QY 354 EVKTSNGLSLSSHEQDYPGNNAKEMKDKKIBKIKAGIMKOYGVKRESIVNKK--- 410
 DB 42 NSLTRHG---OSNEHQDH--DLVEEDDDDDLEPEVIDAPGVTKPRETNVPAEEIIP 96
 QY 411 -NAIYPSGDHHDADPIDHKPV---GIGH---SHS-----NYELFKPEGV 450
 DB 97 PCKTVPVPSDYTKP-TESPVQEAASYGHDAHSHVATFTTSDKEEKRDYPIHHPLSEL 155
 QY 451 AKKENKVTYTGELNVNLLKNSFFNQNT-----LANGOKRVSPSPF----- 495
 DB 156 SDRESETHESLNTVPVSLSGTDTSTFAPSGDDDEYLDGQRKVNVTITLLEESAV 215
 QY 496 -----PELEKLGINKVLITPDOK----- 516
 DB 216 SDYLSGVSNYSQKVTDPKTEETGGVPEIAESFG-NMEYTESPDQKPGQFERDLSTRSKE 274
 QY 517 -----VLEK-----VSGKVGEGVNIANFELDQPLPGOTFKYTIASKDYP 558
 DB 275 FKEFOODFDSVLGKDSPAKFPGESGVVFPVGFSGSGAELE-----KDFP 319
 QY 559 EYSDGTFTVPTSLAYKMASTIFYPFHAGDTYLRVNPQ-FAPVPGTDALVRFDFHGN 617
 DB 320 TRSHDFDMKTEGMDTNSPSRSHSHEPDLKTESGNDKNSPMGFGSESGAE-LEKEFDQKNS 378
 QY 618 AYLENNYKVEIKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQSTIYVEVP1----- 670
 DB 379 G--RNEYS-----PESDGGIGAPLGGNFPVR-SHELDLKNESDIDKDVPTGDFGEPD 427
 QY 671 -----LEKENOTD-----KPSILPQFKRK 690
 DB 428 FLAKRPGTGEAEEEDKFPARSDDEVEVELGDDPTEYLDQFSPSLSPKDERDFESK 487
 QY 691 AQENSKLDERVEEPTSEKVEK-EKLSETGNTSNTSLEEVTVD-----PVOEKVAKFA 744
 DB 488 DDFEETRDENTEPEKQSTYTEKFPASMLGYSGETPVGDTQVAGTVDEKLTVPNEKDOETE 547
 QY 745 ESYGMKL-----ENVLENDGTIELYLPSEGVKKKNAADFTGAPGNGE 789
 DB 548 SAVTTLKPLISGGSGVEEQGEDKSVSGRDIYVAELKUTTEE--DKAFSDNVAEKLOIGE 605
 QY 790 NKPSNGKVTGTVENOPTENKPADSLPEAPNEK-----PVKPE 828
 DB 606 EEKKE---TTTKEVEKISTE-KAASEGEAVEEVRGGGGMVGRIKGFGGGATDEVKPE 661
 QY 829 N-----STDNGMLNPNVGNVSDPMLDPALEAP-----AVDPVQEK 865
 DB 662 SPHSVEEAPKSSGFWGGGATEEVKPKSPHSVEESPSQSLGSTVVPVQOKEL 710

Best Local Similarity 18.6%; Pred. No. 0.94;
Matches 193; Conservative 157; Mismatches 334; Indels 351; Gaps 51;

Y	15	LAKYTTDEG-----YLFDSWIKKSLSEARAA-----AAQYAEKGLT	55
b	701	VAASETTDEGGHSHHTLSQSGETEVEETEVEETVGHHTTTITLPKESAPKVKV	760
Y	56	PPSTDRQDSGTEA-----KGAEIYVNRVKA	83
b	761	ENSIEHKSNDNSOALTKVYLKDLDEFTKSYCHIKYILVSNSSMDQKLLLEVNLTP	820
Y	84	KV-----PLDRMPYNLOQYVVKNGSLIPSYDYHNNKFKFEFDEGLYEA	138
b	821	KELKSCDPLDLL-FNIQ-----NNIPAMTSLYDSMNNDQLHFFE-LYQ	866
Y	139	ATVKYVYERNASDHVRKNKADQSKDPDEKHEDEVSEPTHPESDEKENHAGLNPSADNL	198
b	867	-----YYLHKLKEENHIKLLLEQ-----KQITGTSSTSPGNTTVN-TAQA	908
Y	199	YKPTDTEETEERADTTDEAIEPGTPIRONAMETLTGLKSSLLGLTKDNNNTISAEVDS	258
b	909	THSNSQOQSNASSTNTQNGAVSSGPAYVEESHDLPTVL-----SISNDLKG	956
Y	259	LLALKESQAPIQGOIGOPTLPNNSLATPSPSLPINPGTSHKHEEDGQYGFADNRITA	318
b	957	IVSLL-----NLG-----NKTVPNP-LTIST-TEMEKPYEN-----ILK	989
Y	319	EDESGFVMSHGDNSHYFFKRD-----LTEQIKAAQKHLLEVYKTS-----HN	361
b	990	NNDYF-----NDDIKQPVKSNKVTGLTETQKNALNDEIKKLKDTLOLSFDLYN	1045
Y	362	LDSLSSHEQYDGNKEMKD--LDKKIEKIAIMKQGVKRE-SIVVKNENAI	414
b	1046	LDRLFNKKELGODKMOIKKLTLKQESKLSLNNPHNVQNFVFFNKKEAEIAT	1105
Y	415	-----YPSGDHHAIDIEHKPVGI--GHSNHYELFKPEGVAKK	453
b	1106	ENTLENKILLKHYGLVKNYNGE---SSPLKTLSEVSIQIEDNYANLEKFRALS	1162
Y	454	EKNVYTGEE-----LTVNVLKNSTFNQNFILANGQKRVSEFPPPELKKGIN	505
b	1163	LNDNLHLGKKLSFLSSGLHLITELK-EVINKNYT-----GNSPSENKK--VN	1210
Y	506	MLVKL---ITPDGKVLKSVKGVGNGVNIANPELQPYLPQGTFRYTI-----ASKDY	557
b	1211	EALKSYENFFPEAKYTVTVTP-----PQDVTSPPLSVRVSGSGSTKEE	1255
Y	558	PEVSYDGTFTVPTSLAYKASQTIFFPHAGDTYLRVNPQFAPVPGTD-----AL	607
b	1256	TOIPTSG-----SLLTELQVQVQNTYDEEDSLWLPFGESEDNDEYLDQVVTGE	1309
Y	608	VRYDFEFGNAYLENNYKVEIKLPIPKLNQGTWRTAGNKIP---VTFMANA-----	656
b	1310	SVTMDNILSG--FENEYDVIYK-PL---AGVYSLKKQIEKNIITFNMLNDILNSRL	1362
Y	657	-----YLD-----NQSTYVEVPLEKENQTDKPSILPQFRNK-----	690
b	1363	KKRYFDVLESLDMQFKHISSEYIIE-DSFKLLNSEQKNTLLKSYIKIESVENDIKF	1421
Y	691	AOENSKLDEK-----EETPSKVEKEKLESTGNTSTNLEEYVTPVQEKVAK	742
b	1422	AQEGISYIEKVLAKYKDDLESIKKVEKEKFPSPPTTTPS-----PAKTDEQKESK	1476
Y	743	FA-----ESYGMKLENVLFNN-----DGTIELY	765
b	1477	FLPFLNIETLYNNLYNKIDYLLINKAKINDCNVKEDEARVKITKLSDLKADIDKIDLF	1536
Y	766	LPSGEVKKMADFTGEAPQNGENKPSNGK-VSGTGVENQPT-----ENKPADSLP	817
b	1537	-----KNTNDEFAIKKLINDOTKDKMLGLKLLSTGLVONFPNTIISKLEGKQDMLN	1588
Y	818	EAPNE--KPVKPPENS	830

MSPI1_PLAFM STANDARD; PRT; 1701 AA.

AC P08569; 01-AUG-1988 (Rel. 08, Created)

DT 01-MAY-2000 (Rel. 39, Last sequence update)

DT 01-MAY-2000 (Rel. 39, Last annotation update)

DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (P190).

DE (PMSA) (P190).

GN MSP-1.

OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_Taxid-70153;

RN [1]

RN SEQUENCE FROM N.A.

RX MEDLINE-89011243; PubMed-3079521;

RA Tanabe K., Mackay M., Goman M., Scaife J.G.;

RT "Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum."

RT Plasmodium falciparum.

RL J. Mol. Biol. 195:273-287(1987).

RN [2]

RN REVISIONS TO 1403; 1569 AND 1629.

RA Tanabe K.;

RP Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RN SEQUENCE OF 1-115 FROM N.A.

RX MEDLINE-86136024; PubMed-3004972;

RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,

RA Stunnenberg H., Bujard H.;

RT "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level."

RT Plasmodium falciparum merozoites.

RL EMBO J. 4:3823-3829(1985).

CC [1-] SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).

CC [1-] PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC [1-] THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X05624; CAA29112.1; -

DR PIR; A26868; A26868.

DR PIR; B25120; B25120.

DR InterPro; IPR000561; EGF-like.

DR Pfam; PF00008; EGF; 1.

DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.

KW SIGNAL

KW CHAIN

FT SIGNAL 1 19

FT CHAIN 20 1701

FT CARBOHYD 110 110

FT CARBOHYD 239 239

FT CARBOHYD 470 470

FT CARBOHYD 536 536

FT CARBOHYD 607 607

FT CARBOHYD 802 802

FT CARBOHYD 899 899

FT CARBOHYD 919 919

FT CARBOHYD 965 965

FT CARBOHYD 991 991

FT CARBOHYD 1089 1089

FT CARBOHYD 1196 1196

FT CARBOHYD 1588 1588

FT SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 3.6%; Score 172.5; DB 1; Length 1701;

us-09-884-465a-332.isp

Fri Mar 28 09:21:11 2003

Db 1589 ISQHCVKQCPENS 1603

Search completed: March 27, 2003, 11:55:11
Job time : 44 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:53:13 ; Search time 103 seconds
(without alignments)
1812.414 Million cell updates/sec

Title: US-09-884-465A-332

Perfect score: 4728

Sequence: 1 MQITYTDEIQVAKLQVY.....IELRPFGEVKKNLSDFA 906

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3357	71.0	1039	16 Q9ANY1	Q9any1 streptococc
2	1307.5	27.7	839	16 Q9ANY2	Q9any2 streptococc
3	946.5	20.0	819	16 Q97QM9	Q97qm9 streptococc
4	946	20.0	844	2 Q9AG74	Q9ag74 streptococc
5	935.5	19.8	819	2 Q9ANY3	Q9any3 streptococc
6	636	13.5	802	16 Q97QM8	Q97qm8 streptococc
7	636	13.5	816	2 Q9AHY9	Q9ahy9 streptococc
8	233	4.9	825	2 Q9JG75	Q9jg75 streptococc
9	230	4.9	825	16 Q99XV4	Q99xv4 streptococc
10	223.5	4.7	822	2 Q9ZHG7	Q9zhg7 streptococc
11	215.5	4.6	1524	10 Q8RYN2	Q8ryn2 oryza sativ
12	213.5	4.5	5458	5 Q9U459	Q9u459 plasmodium
13	208	4.4	2647	5 Q9U4X0	Q9u4x0 plasmodium
14	202.5	4.3	2269	5 Q96223	Q96223 plasmodium
15	201.5	4.3	1078	5 Q963T1	Q963t1 plasmodium
16	201.5	4.3	1873	2 Q9Z4N7	Q9z4n7 enterococcu

17	201	4.3	1790	3 Q07380	Q07380 saccharomyc
18	201	4.3	2004	16 Q97QP7	Q97qp7 streptococc
19	198	4.2	1420	5 Q9Y0V8	Q9y0v8 drosophila
20	197	4.2	3484	5 P91257	P91257 caenorhabdi
21	195.5	4.1	1139	5 Q97237	Q97237 plasmodium
22	195	4.1	1698	2 Q9LC00	Q9lc00 staphylococ
23	195	4.1	2910	10 Q9FND5	Q9fnd5 arabidopsia
24	194.5	4.1	1134	2 Q99051	Q99051 streptococc
25	193.5	4.1	1125	16 Q98PM9	Q98pm9 mycoplasma
26	193	4.1	1245	2 Q9L470	Q9l470 staphylococ
27	192	4.1	1871	10 Q9SRD5	Q9srd5 arabidopsia
28	191	4.0	3111	5 Q9VH10	Q9vh10 drosophila
29	190.5	4.0	2747	5 Q9BJX9	Q9bjx9 plasmodium
30	190	4.0	1596	11 Q62766	Q62766 rattus norv
31	190	4.0	4688	16 Q9PQ08	Q9pq08 ureaplasma
32	189.5	4.0	2748	3 Q9VC00	Q9vc00 drosophila
33	189.5	4.0	2768	5 Q9VC00	Q9vc00 drosophila
34	189.5	4.0	5005	16 Q9PP25	Q9pp25 ureaplasma
35	189.5	4.0	5327	5 Q76891	Q76891 drosophila
36	189	4.0	1596	11 Q9Z1F7	Q9z1f7 rattus norv
37	188.5	4.0	1793	5 Q9W596	Q9w596 drosophila
38	187	4.0	2151	5 Q9NG79	Q9ng79 trichomonas
39	185.5	3.9	1983	16 Q9CU55	Q9cj55 lactococcus
40	185.5	3.9	2735	5 Q76416	Q76416 caenorhabdi
41	185	3.9	1733	2 Q9KI14	Q9ki14 staphylococ
42	185	3.9	2485	5 Q96134	Q96134 plasmodium
43	184.5	3.9	1964	2 Q59947	Q59947 streptococc
44	183.5	3.9	1795	2 Q9LC39	Q9lc39 staphylococ
45	183	3.9	865	2 Q9FDJ1	Q9fdj1 streptococc

ALIGNMENTS

RESULT 1

ID	Q9ANY1	PRELIMINARY;	PRT; 1039 AA.
AC	Q9ANY1;		
DT	01-JUN-2001 (TREMblrel. 17, Created)		
DT	01-JUN-2001 (TREMblrel. 17, Last sequence update)		
DT	01-OCT-2001 (TREMblrel. 18, Last annotation update)		
DE	Pneumococcal histidine triad protein E precursor (Hypothetical protein SPI004).		
GN	PHTE OR SPI004.		
OS	Streptococcus pneumoniae.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;		
OC	Streptococcaceae; Streptococcus.		
OX	NCBI_TaxID=1313;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=21101045; PubMed=11159990;		
RA	Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,		
RA	Dormitzer M., Degan R., Brewah Y.A., Warren P., Lathigra R.,		
RA	Langermann S., Koenig S., Johnson S.;		
RT	"Identification and characterization of a novel family of pneumococcal proteins (the Pht family) that are protective against sepsis."		
RL	Infect. Immun. 69:949-958(2001).		
[2]			
RN	SEQUENCE FROM N.A.		
RP	STRAIN=TIGR4;		
RP	MEDLINE=21357209; PubMed=11463916;		
RA	Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,		
RA	Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,		
RA	Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,		
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,		
RA	Holtzapple E., Khouri H., Wolf A.M., Otterback T.R., Hansen C.L.,		
RA	McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,		
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,		
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,		
RT	"Complete genome sequence of a virulent isolate of Streptococcus pneumoniae."		
RL	Science 293:498-506(2001).		
DR	EMBL; AF318956; AAK06761.1; -		

Db 818 TISAEVDSLLALKESQAPAIQ 839

RESULT 3

Q97QM9 PRELIMINARY; PRT; 819 AA.

AC Q97QM9

DT 01-OCT-2001 (TREMBlrel. 18, Created)

DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Conserved domain protein.

SP1174.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TIGR4;

RX MEDLINE=21357209; PubMed=11463916;

RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,

RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,

RA Durkin A.S., Gwion M., Kolonay J.F., Nelson W.C., Peterson J.D.,

RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,

RA Holtzapple E., Khouri H., Wolf A.M., Otterback T.R., Hansen C.L.,

RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,

RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,

RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;

RT "Complete genome sequence of a virulent isolate of Streptococcus

RT pneumoniae";

RL Science 293:498-506(2001).

DR EMBL; AE007418; AAK75283.1; -

KW TIGR; SP1174; -

KR Complete proteome.

SQ SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;

Query Match 20.0%; Score 946.5; DB 16; Length 819;

Best Local Similarity 63.8%; Pred. No. 9.7e-38;

Matches 199; Conservative 23; Mismatches 41; Indels 49; Gaps 7;

QY 2 QIYTTDEIQVAKLAGKYTTEDGYIFD-----TSWIKKDSLEAER 42

DB 511 QIYTTDEIQVAKLAGKYTTEDGYIFDPRDITSDGDAYVTPMTHSHWIKKDSLEAER 570

QY 43 AAQAQAYAKEGLTPPSTDHQDSGNTKAGAEIYNRVKAAKVPDLRMPYNLYQYTVEN 102

DB 571 AAQAQAYAKEGLTPPSTDHQDSGNTKAGAEIYNRVKAAKVPDLRMPYNLYQYTVEN 630

QY 103 GSLIIPSYDHYHNKFEWDEGLYAPKGYSLDGLATVYVE-----PR-----NAS 151

DB 631 GSLIIPSYDHYHNKFEWDEGLYAPKGYSLDGLATVYVEHNPENRPHSDNGFGNAS 690

QY 152 DHVRKKNK-----ADQDSKPEDKEHDEVSFTHPESEKENHAGLN-----SADNL 198

DB 691 DHVRKKNQADNTQTEKPEKQTEKPEETPREKPEKPEKPEKPEKPEKPEKPEKPEK 750

QY 199 YKSTDTTEETEE---EADTDTDEAIPGTPSIRONAMETLTGLKSSLLITKNNWISAE 255

DB 751 EEPQVETERYEKLREAEALGLKIQI-----DPIKSNKAKETLTGLKNNLLFGTDQNTIMAE 807

QY 256 VDSLALKESQ 267

DB 808 AKLLALKESK 819

RESULT 4

Q9AG74 PRELIMINARY; PRT; 844 AA.

AC Q9AG74

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE PpA.

GN PpA.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21246685; PubMed=11349048;

RA Zhang Y., Masi A.W., Barniak V., Mountzouras K., Hostetter M.K.,

RA Green B.A.;

RT "Recombinant PpA Protein, a Unique Histidine Motif-Containing Protein

RT from Streptococcus pneumoniae, Protects Mice against Intranasal

RT Pneumococcal Challenge";

RL Infect. Immun. 69:3827-3836(2001).

DR EMBL; AF340221; AAK26629.1; -

SQ SEQUENCE 844 AA; 94769 MW; D738A55290FF8902 CRC64;

Query Match 20.0%; Score 946; DB 2; Length 844;

Best Local Similarity 62.5%; Pred. No. 1.1e-37;

Matches 202; Conservative 22; Mismatches 33; Indels 56; Gaps 8;

QY 2 QIYTTDEIQVAKLAGKYTTEDGYIFD-----TSWIKKDSLEAER 42

DB 531 QIYTTDEIQVAKLAGKYTTEDGYIFDPRDITSDGDAYVTPMTHSHWIKKDSLEAER 590

QY 43 AAQAQAYAKEGLTPPSTDHQDSGNTKAGAEIYNRVKAAKVPDLRMPYNLYQYTVEN 102

DB 591 AAQAQAYAKEGLTPPSTDHQDSGNTKAGAEIYNRVKAAKVPDLRMPYNLYQYTVEN 650

QY 103 GSLIIPSYDHYHNKFEWDEGLYAPKGYSLDGLATVYVE-----PR-----NAS 151

DB 651 GSLIIPSYDHYHNKFEWDEGLYAPKGYSLDGLATVYVEHNPENRPHSDNGFGNAS 710

QY 152 DHVRKKNK-----ADQDSKPEDKEHDEVSFTHPESEKENHAGLNPSADNLYKPTDTE 206

DB 711 DHVRKKNQADNTQTEKPEKQTEKPEETPREKPEKPEKPEKPEKPEKPEKPEKPEK 761

QY 207 ETEEA-EDTDEAIP-----GTPSIRONAMETLTGLKSSLL 244

DB 762 ETEESPESPESEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESE 821

QY 245 GTRDNNTISAEVDSLLALKESQ 267

DB 822 GTQDNNTIMAEAKLLALKESK 844

RESULT 5

Q9ANY3 PRELIMINARY; PRT; 819 AA.

AC Q9ANY3

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE Pneumococcal histidine triad protein B precursor (fragment).

GN PHTB.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21101045; PubMed=11159990;

RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,

RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,

RA Langermann S., Koenig S., Johnson S.;

RT "Identification and characterization of a novel family of pneumococcal

RT proteins (the Pht family) that are protective against sepsis";

RL Infect. Immun. 69:949-958(2001).

DR EMBL; AF318954; AAK06759.1; -

KW Signal.

FT SIGNAL

FT NON_TER

1 29 POTENTIAL.

819 819

```

Db      573 VAAQAYTKEGILPPSPDADVANKPAGTSDSAAAIYNNRVKGEKRIPLVRLPYMVEHTVEVKN 632
QY      103 GSLIIPSYDHYHNKIFWFDGELYEAPKGYSLDGLIATVKYVE-----PR-----NAS 151
Db      633 GNLIIPKDHYNKIFAWFDQTYKAPNGYTLDELATIKYIYVEHDPDRPHSDNGWGNAS 692
QY      152 DHVRKNKADQDSKPEDKEHDEVSEPTHPESDEKENHAGLNPSADNLKXPSTDTEETEE 211
Db      693 EHV-LGKKHSDPNKNFKADE--EPVEETPAPE-----VPQVETEKVEAQ 736
QY      212 AEDTTDEAEI----PCTPSIRQNMETLTGLKSSLLGTCKDNTTISAEVDSLLALKESQ 267
Db      737 LK-----EAEVLLAKYVDSLSKANATETLAGLRNLTQIOMDNNISMAEAKLLALLKGSN 792
QY      268 PAPIQGPOI 276
Db      793 PSSVSKEKI 801

RESULT 7
Q9AHT9
ID Q9AHT9 PRELIMINARY; PRT; 816 AA.
AC Q9AHT9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Pneumococcal histidine triad A protein.
GN PHTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
RX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RC MEDLINE=21116976; PubMed=11179332;
RA Wizemann T.M., Helmricks J.H., Adamou J.E., Erwin A.L., Kunsch C.,
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
RA Langermann S., Johnson S., Koenig S.;
RT "Use of a Whole Genome Approach To Identify Vaccine Molecules
RT Affording Protection against Streptococcus pneumoniae Infection.";
RL Infect. Immun. 69:1593-1598(2001).
DR EMBL: AF291695; RAK19155.1;
SQ SEQUENCE 816 AA; 91519 MW; 5359126A611D27ED CRC64;

Query Match 13.5%; Score 636; DB 2; Length 816;
Best Local Similarity 45.3%; Pred. No. 8.7e-23;
Matches 140; Conservative 40; Mismatches 75; Indels 54; Gaps

QY      2 QIYTTDDEITQVAKLACKYITEGYIFD-----TSWIKKDSLEAER 42
Db      527 QIYETEDEVRIQLADKYITSDGYTFDEHDIISDEGDAYTTPMGRSHHWIGKDSLSDEK 586
QY      43 AAAQAYAKEKGLTPPSTDHODSNTAEKAGAEALYNNRVKAAKVPIDRNPYNLQYTVEVKN 102
Db      587 VAAQAYTKEGILPPSPDADVANKPAGTSDSAAAIYNNRVKGEKRIPLVRLPYMVEHTVEVKN 646
QY      103 GSLIIPSYDHYHNKIFWFDGELYEAPKGYSLDGLIATVKYVE-----PR-----NAS 151
Db      647 GNLIIPKDHYNKIFAWFDQTYKAPNGYTLDELATIKYIYVEHDPDRPHSDNGWGNAS 706
QY      152 DHVRKNKADQDSKPEDKEHDEVSEPTHPESDEKENHAGLNPSADNLKXPSTDTEETEE 211
Db      707 EHV-LGKKHSDPNKNFKADE--EPVEETPAPE-----VPQVETEKVEAQ 750
QY      212 AEDTTDEAEI----PCTPSIRQNMETLTGLKSSLLGTCKDNTTISAEVDSLLALKESQ 267
Db      751 LK-----EAEVLLAKYVDSLSKANATETLAGLRNLTQIOMDNNISMAEAKLLALLKGSN 806
QY      268 PAPIQGPOI 276

```


Query Match 4.78; Score 223.5; DB 2; Length 822;
Best Local Similarity 19.68; Pred. No. 0.0065;
Matches 186; Conservative 143; Mismatches 319; Indels 301; Gaps 52;

RESULT 11

Q8RYN2	Q8RYN2	PRELIMINARY;	PRO; 1524 AA.
ID	Q8RYN2	AC	
DT	01-JUN-2002	(TRENBlrel. 21, Created)	
DT	01-JUN-2002	(TRENBlrel. 21, Last sequence update)	
DT	01-JUN-2002	(TRENBlrel. 21, Last annotation update)	

DE P0663E10.13 protein.
GN P0663E10.13.
OS Oryza sativa (japonica cultivar-group). Embryophyta; Tracheophyta;
OC Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eubacteriobacteriota; Bacteriota; Proteobacteria; Gammaproteobacteria; Rhodospirillales; Rhodospirillaceae; Rhodospirillum rubrum
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RL clone: P0663E10."
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
EMBL: AF004317; BAB90730.1; BAF90730.1;
SQ SEQUENCE 1524 AA; 168701 MW; E47FE1A9C5842228 CRC64;

Query Match 4.6%; Score 215.5; DB 10; Length 1524;
Best Local Similarity 18.9%; Pred. No. 0.036;
Matches 189; Conservative 139; Mismatches 371; Indels 299; Gaps 42;

QY 56 PSTDQDQSGNTEAKG-----AEIYRVKAAKVPDLPMPYNLOY 96
DB 72 PPAARDGAGAGAYGEIFRFDGFAEPYEDLLAEVALAAEIASSESSRS-----125
QY 97 TVEVKNGL-LIPS--YDVHNIKF-EWFDGLY-----EAPKGYSLDILLAT-----140
DB 126 SVKESGOLDADPILHGHVSTVGYDQHDDEFEFSISPPDPSGKOFMSYNKATGRPD 185
QY 141 -----VKYVPEPNASDRVRNKAQDQSKPDEKHEDEVSEP-----THPES 182
DB 186 DIVKMTTCVPEPISYVDSRNI-----NKSAMDQVVVVDCDTPFANGKSGNGLFPSS 240
QY 183 DEKENHAGLNPADNLYKP-----STDTEETEEAEDTTDEAIEPGTFS-----IRQ 229
DB 241 SLLKSSASSDVAQDLHTPICHPIKNDCEDEYHKLSTHSASSEVPSPDYPPFLRVSN 300
QY 230 NAMET-----LTGLKSLILGTDKNNTISABVDSLLALKKESQAPITGQP 275
DB 301 NSLHTPIKVPOLLAPSKLLNKKESKANGKSTGLTFPSSS-----SVKSSADPMADQN 357
QY 276 IGOPT-----LPNNSLAT---PSPSLPINPGTSHKHEBEDYGFDAIR 315
DB 358 LHPTCHPIKTCDEDEYHKLSTHSASSEDPSPDPLRVNPNLSLHTPIKVPQPPSK 417
QY 316 IIADESGFVMSHGDSN-----HYFFKK--DLTEQIKAAQKHLKEVTS-----358
DB 418 LLAKKES---KANGDSVSTNSAAAAAIAKEAMFEARLKAALKEAMERKGDSPKLRKKP 474
QY 359 --HNGLDLSLSSHQDYPGNAKEMKDLKDKTEKIKIAGIMQYGVKRESIVYVK--EKNAIL 414
DB 475 GHNGYKSTELKSMAP---EVRVYDEKL--TWRRIVKEEYEEYALVKNKGDSSAVN 529
QY 415 YPGDHHADPIDHKKVPVIGHSHN-----YELFKPEGVAKKGNKRYTG 461
DB 530 LTHCDHNEKGVLPKQPHQAQSGKLEQLGKWTSGAEFTVLISPDQ---KCKTNSV---583
QY 462 EELTNVYLLKNSTFNQNTLANGOKRVSFSPPELEKLGINMLVKLITPDGKVLKV 521
DB 584 -----TCGDNVQTNPPSKLG-----QPEKKG-----ETT 610
QY 522 SKGVFGGV---GNIANFELDQPLPGQTFKTYIASKD-----YPEVSYDGTFTVPT 570
DB 611 SGDPVGGKSWDGDIAELRMEHVNL---EYALGSTDGCKAPTAPESFSN--EKPT 664
QY 571 SLAYKMASQITFYFPFHAGDYLRVNPQFVPGKTDALRVDFEPHGNAYLNNYKVEGK 630
DB 665 ---YQESTETHFKECVGAQNT-----QERYGDGGAFTSCVDSK 701
QY 631 LPPLKNGQITRTA-----GNKIP-----VTFMANAYLDNQSYIYVEVPLEKENQ 676
DB 702 LHAPEIFCASLSCISGHCNGKNSPDSASTKETTSLGSKENKNNTEALEVPCAD-EMQ 760

QY 677 TDRPSILPQPKRNKAQEN-----SKLDEKVEBPKT-----SERVEKEKL- 715
DB 761 SOILQYHEPRNENIDKKAQVKSLEESVEYETPNFQKSSSTAHTGETETVEKQMF 820
QY 716 ---SETGNSTNSLEEVPTVPQEKVAKFA--ESYGMKLENVLFNMDGTIE-----763
DB 821 SFSDELRPQNKIGITEAPPELSLHKEIKKFTGEKAVITLEGDVVQKSGSLREANITL 880
QY 764 -----LYLPSGEVKKNNMADFTGAPOQNGENKPKSVKSTGVTEVNOPT 808
DB 881 ESASANEAEAEANAFVEGINVMEHSTY--GTSVEDSDQIDSENRMDCMGDLVSHGN 939
QY 809 ENKPADSLPEAPNPKVPKPNSTNGMLNPNCGVSDP 846
DB 940 EEAADPWLDNSEKQVEEIFSHEEGQLSVREGIDGP 977

RESULT 12
QY 90459 PRELIMINARY; PRT; 5458 AA.
AC Q90459; Q9GPG4;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
DE Erythrocyte membrane-associated giant protein antigen 332.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCCL/HN;
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M., Lu J.H., X J.;
RT "Molecular cloning and structure analysis of the plasmodium falciparum
erythrocyte membrane-associated giant protein Ag332 (Pf332) gene."
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
EMBL: AF202180; AAF15293.3;
DR InterPro: IPR001313; Pumilio/Puf.
SQ SEQUENCE 5458 AA; 615269 MW; 4F6A37D92BD80172 CRC64;

Query Match 4.5%; Score 213.5; DB 5; Length 5458;
Best Local Similarity 19.2%; Pred. No. 0.26;
Matches 197; Conservative 181; Mismatches 419; Indels 227; Gaps 44;

QY 6 TDEIQVAK-LAGKTYTDEGYIFDTSWIKDLSLSE---AERAAQAYAKEGLTPPSTPH 61
DB 1114 SDEIVVEGSGATEIIVVEGSGATEVV-VVEGVSVEEMLEEGSGATEVVEEGL---SSDN 1169
QY 62 QDSGNTGAEAGAIYRVKAAK-----KVPLDRMPY-----NLQYTVKNGS 104
DB 1170 VOKSGVIENVEGIYS-VTKADESNKNIPLKESFVDESGFKGGPTDNVSDVDNSE 1228
QY 105 LIIPSYDHYHNKFEWDEGLYEAQKYSLEDLLAT-----VKYVPEPNASDRVRNKA 159
DB 1229 DIINEHTPELEKIE-----ELPTEYTTADIHTRGETEYKNIYKINEEVEKAF 1281
QY 160 DQDSKPD-----EDKEHDEVSEPT--HPSEDEKENAGLNPADNLYKFTDTEETEEBA 212
DB 1282 QEEKITENIPVERSEVTEIDVQPSLAQVEQKE-----SDTNEIEE-- 1323
QY 213 EDTTDEAEIPGTPSIQONAMETITGLKSLILGTDKNNT---ISAEDVSLALLKESOPA 269
DB 1324 TKLANEKIIPV-SVTENVVEK--EGLDTEVLEDESEITEEIVEEVSACEEIVEE--- 1379
QY 270 PIQGPQIGOPTLPNNSLATPSPSLPINPGTSHKHEBEDYGFDAIRIADSESGF--VMS 327
DB 1380 -----SSSEIVEE--ESSEIVEEESSEIVEE 1409
QY 328 HGDSNHYFFKADTEPQIKAAQKHLKEVTSNGLSLSSHEQDY-----PGNAKEMKL- 382
DB 1410 EESS---SEEIVEEESSEIVEEESVTEVLEEGSVTEEIVEEVSSTTEEVKDIG 1465
QY 383 ---DKKIEKIAK-----IMQYGVKRESIVVKNKNAIYPSGDH-----HHADPIDE-- 428

```

/SEVLEEBEGFGTEEFVQGGSGVIEIVETESTERKVLVEDVGNSEIIVQEBGPVQAEI 1525
49 -HKPVGIGHSHNTLFRKPGVAKKGN-----KYTGTELTNVNL----- 470
1576 VHEEVSITTEKDEVDIRSTEE-IVKVGVSSEIIVESVASEEIVVEGVSVEEVBEEK 1584
QY 471 LKNTFNQNTLANGKRVSPFPPELEKLG-----INMLVLIIPDCKVL 518
Db 1585 LINEVGETESTEIVQKEVS-----DAEVLQGGSGNNEEILKESIVEIVGPEGSVT 1639
QY 519 EKVSQVPGEGVGNIANFELQPLPGQTFYTIASKDIPYVSVDGTFVPTSLAYKMAS 578
Db 1640 EBI-----VDYGSFAEVEEELVTEAVQEGSVTE--BIKEESITENAEESASA 1691
QY 579 QTIFYPFHAGDTYLRVNPQFVAPKGTDA-----LVVPDEFHGNAYLNNKYGE----I 629
Db 1692 ETI-----EEKGN-----NDEIVKEGLDTEIIVNEVSVTDEVEEKLNEQIVGEERSVT 1744
QY 630 KLPKPLNAGTTRTAGNKIPVTENANAYLDNOSTVI-----VEVPTLEKENTDKPSILPQ 685
Db 1745 EKPVEVERSATDELVEEGASVTEKVSVEHGGSTEQILDESVAEIVEEVSVDKIIIEE 1804
QY 686 FRNKA--QENSKLDEKVEEPK-----TSKVEKEKLSGTGNSNSTLEEV 730
Db 1805 VGVDEVEEGSVIEEIVEEESVPEEILEELSSEVLEDEWVTDAFMGQEGSVIEEI 1864
QY 731 PV-----DPVQSKVA-----KFAESYGMKLEWLFNMDGTIELYPSGEVIR-----K 774
Db 1865 EBIVDGEGSITEIIEVDGSAKEIIEEVSFVYELGREGFVIEEIEEGSVIEQVEDTK 1924
QY 775 NNADFTGAPOGNGENKPSGKSTGTVENOPTENKPADSLPEAPNKPVPKPNSTONG 834
Db 1925 TVSEKSEESAETEEVKEVEKESIESEKIVEKEESTVEIVRQEESTEKIVKDVSPTDF 1984
QY 835 MLNPEGVSGDPMLDPALEAPVDPVQEKL-EKFTASYGLDLSVIFNMGTIELRPS 893
Db 1985 VEQ-----TDSVTEKVIQEGSNTSEVABDEVEEKESASDEHEQEDVSNAQVTCCKSVT 2038
QY 894 GEVI 897
Db 2039 KEIV 2042

RESULT 13
Q904X0 PRELIMINARY; PRT: 2647 AA.
AC Q904X0;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DE 01-MAR-2002 (TREMELrel. 20, Last annotation update)
DE Putative erythrocyte binding protein EBL-1 (Fragment).
GN EBL-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP MEDLINE FROM N.A.
RX MEDLINE=20078864; PubMed=10613703;
RA Peterson D.S., Wellens T.E.;
RT "EBL-1, a putative erythrocyte binding protein of Plasmodium
RT falciparum, maps within a favored linkage group in two genetic
RT crosses."
RL Mol. Biochem. Parasitol. 105:105-113(2000).
DR EMBL; AF131999; AND33018.1;
FT NON_TER 2647 2647
SQ SEQUENCE 2647 AA; 304550 MW; AB98F88FD754E300 CRC64;

Query Match 4.4%; Score 208; DB 5; Length 2647;
Best Local Similarity 19.9%; Pred. No. 0.18;
Matches 201; Conservative 151; Mismatches 384; Indels 274; Gaps 48;

QY 58 STDHQDSGN-----TEAKGAEATYNNRKAAC-----KVPIDRMPYNLYQTYVE-- 99
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |

```

```

Db 1243 SFDHVDSPNISELSQASONADS-YQCEKPSKNILRTDGVSEKEKISVSPNPNVSVTYDBG 1301
QY 100 -----VKNGLIIPSDYHYNHIFWF-----DEGLYAPAGYSLEDLL----- 138
Db 1302 DKRGISDDSSIHEDPEKNLHYESFISEGGLGEDIEKERKEDGSLTPSPKINDP 1361
QY 139 ATKVYVPEPNADHYKKNKADQDQSKPDEKDEHVESEPTHE-----SDE--KENHAGLN 192
Db 1362 GKRENIYDP-SVERVDDNRSDISNVVSEQESSIISSRNGTEGINNSEELKSEETSD 1420
QY 193 PS-----ADNLYKSTDETEEEAE-----DTDEAEIPGT-----PSIRNANMETLT 236
Db 1421 VNNRDEHNKQNLVSSSTQOESERERKENADSSHELSISEVGETIRRDAAEASE 1480
QY 237 GLKSSLLGFKDNNTISAEVDSLALLKESQAPAQOQOIGOPTLPNNSLATSPSLPIN 296
Db 1481 NDKGEDIQSEEQOIVTEKI-----KLIETV--NIL---QPSTPL- 1516
QY 297 PGTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDSNHYFF-----KKDLTEEQI 345
Db 1517 -GEIHKKESEE-----IDDKONTGCVTHTDANTFOYRSSESVEVITNKPDEMNTT 1567
QY 346 KAAQKHLEEVKTSINGLDSLSHEDQYPGNAKEMKOLDKKEEKIAGIMQYGVKRESIV 405
Db 1568 KPPSOYIEK---EIGEDSTKNQNDPOSNS-IIPLINKNEEDGVSIPTSRNVHESGSP 1623
QY 406 VNKEKNALIYPSGGHHAD-----PIDEHKPVGIGHS--HSNYELFKP 446
Db 1624 VSRNEQ-IIEKDKKHITDDTTINPSENGLKGCEMNDISKVTITITESPLRDVEQMIEP 1682
QY 447 BEGVAKEGKNVYTGEBELTNVNLKSTFNNQF-----TLANGQKRVSFSPPELEKK 501
Db 1683 IDGKGNKNNIIGEPQESTTEIRKQMDGPISNVNIPELHPVAEGSKL-----EAKE 1735
QY 502 LGINMLVK-LTTPGKVLKYSKGVGE-----GVGNIANFE-L 538
Db 1736 RSMDDAGTITEDITVVEDPFG--IGEHQNLKEVHEQASELNTYNSLDGTYNVEVERL 1793
QY 539 DQYLPQGTFRYTIASKDIPYVSVDGTFVPTSLAYKMASOTIFYPHAGTYLRVNPQF 598
Db 1794 DEN--PGSIPNDRITT-EHIELDKEKEIHEPNELDAHNGEQEEMIRNEVSDN--RMDQI 1848
QY 599 AVPKGTALVRVF-----DEFHGNAYLEN-----NYKVGEEKLP----- 632
Db 1849 SRDNETROLNQDHESDSKDEIDKREMLEENPNSSSDSLENPEGKEKGIETHSSEE 1908
QY 633 IPKLNQGTTRTAGNKIPVTENANAYLDNOSTVIIVVEVPTLEKENTDKPSILPQFNRKQA 692
Db 1909 LDSVDEETIKYKGIENQITETIEESVEQODT---NIPGNSKETEDV-----NSRAD 1956
QY 693 ENSKLDERVEEPTSEKVEKELSTGNSNSTLEEVPTVDPVQEKVAKFAESYGMKLE 752
Db 1957 MEERKDVAIKQIITEESEEELEISK-DTTTSHS---EKPSIEEQSVNIVDSKNEINVQIE 2012
QY 753 NVLFN--MDGTIELYPSGEVTKKMAADFTGEAPOGNGENKPE-----NG 796
Db 2013 KNVQNEQNEGDPIILL---EQGNKIAILEN---QKNEYNNPQSLSHKERTILLEVDLEG 2066
QY 797 KYSTGTVENOPTENKPADS---LPEAPN-----EKVPKPNSTONG 834
Db 2067 SMDTNCILSLFNKNK-CDSIQIIPASNTDNKLDITENKDDFSEIEKSVEGIEHENGKD 2125
QY 835 MLNPEGVSGDPMLDPALEAPVDPVQEKLEKFTASYGLDLSVIFNM 884
Db 2126 LLNKESAESDDV-----PVQNKIEHDSNAGV-IDQYEYRTD 2161

RESULT 14
Q26223 PRELIMINARY; PRT: 2269 AA.
AC Q26223;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)

```

Db	678	EGIKKKQOIIVTKDK-----KNIYEEINKLLSEISKEIKONTSLERKVKDINLS-YG	729
Qy	720	NSTNSLTLEEPTVPQVOKVAKFAESYGMLENVLFNMDGTIELYLPSEVTKKKNADF	779
Db	730	QNLGNLFLEQIDEEKKAENTIKSMEAYIDDLNIK-KKQEIETEMDKMDINKMEAL	788
Qy	780	TGEAPQNGENKPSGKSVGTGVENQPT-----ENKPADSLPEAPNEKP	824
Db	789	KISHDDKKCHDKSKHNKENTISDIYKSSKIQDFSRSDINDLNKLQKNVSEQN---	845
Qy	825	VRPENSTNGMLNPGNVGSDPMLDPALEAPVDPVQEKLEKFTASTYGLGL----	881
Db	846	---HNSDINOCLNEVANIYNILKLNKIKKIIDKVKEYTSETEKNKNKINDELNNSEKVIK	902
Qy	882	NMDGTIEL-----LPSGEVKKNL	901
Db	903	KIEGDLSEKRSKINSTLDDXIDCEIKNINVLAKNI	940
RESULT 15			
Qy	Q963T1	PRELIMINARY; PRT; 1078 AA.	
AD	Q963T1		
IC	Q1-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DT	Glutamate-rich protein (Fragment).		
GN	GLURP.		
OS	Plasmodium reichenowi.		
OS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
ON	NCBI_TaxID=5854;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RK	MEDLINE-21313688; PubMed-11420113;		
RA	Theisen M., Thomas A.W., Jepsen S.;		
RT	"Cloning, nucleotide sequencing and analysis of the gene encoding the		
RT	glutamate-rich protein (GLURP) from Plasmodium reichenowi.;"		
RL	Mol. Biochem. Parasitol. 115:269-273(2001).		
RL	EMBL: AF356828; AAK40236.1; ;		
FT	NON_TER 1		
FT	NON_TER 1078 1078		
FT	NON_TER 1078 AA; 123581 MW; 8D388D808223913C CRC64;		
SEQ			
Query Match 4.3%; Score 201.5; DB 5; Length 1078;			
Best Local Similarity 19.9%; Pred. No. 0.11;			
Matches 178; Conservative 152; Mismatches 357; Indels 209; Gaps			
Qy	11	QVAKLAKYTTEDGYFDTGWIKKDSLEAERAAQAYAKEGLTPPSTDHQDSGNTAK 70	
Db	312	EIFSLDKETTNEDILPNQPLENIKQSEIENVQDHLPKENIIDLKLDNEKHIDESQ 371	
Qy	71	GAEAIY--NRYKAARKYPLDORPYNLYQTYEVKN--GSLIIPSYDHYHNKIFWFEGLY 126	
Db	372	HNVLQENINNHOLEPQEKA--NIE-SPEKNIIDSEILPENVEKEEIVD-----V 422	
Qy	127	EAPGYSLDILATVKKYVVEPRNADSVHRNKADQDSKPDDEHDS-VSEPTHPSEDK 185	
Db	433	PSPHFNHETL-----EQTSSEH---EEAVSEKNAHETVEHEEAVSQSNPEKADN 472	
Qy	186	ENHAGLNPADNLKPSDTDETEEEAEDTTDEAIEFPTPSIRONAMETITGLKSLILG 245	
Db	473	DGNYSQNSNNELNENEFVESEKSEHAENES-----SLEAHOEIV----- 516	
Qy	246	TRONTWISAE---VDSLALAKESQAPIQGPQIGPTLPNLSLATPSLPINPGTSHE 302	
Db	517	PEQNNQSGESKLVNDREGGFARHAQEIVEPQ-----NQSGESKLVNDDEGGFEE 568	
Qy	303	KHEED-----GYGPDANRIITAEDESGVMVSHGDS-----NHYFFKDLT --- 344	
Db	570	AHQEIVEPQNNQSGESKLVNDDEGGFEAHENFSEESNELHEHEEVESDESQPEP 629	
Qy	342	-----EEOITAAQKHLEVKTSHNGLSLSDSHEDQYPCNAKEMKOLDKKEEIKAG 397	

Search completed: March 27, 2003, 11:56:59
Job time : 118 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:57:08 ; Search time 20 Seconds
(without alignments)
2660.182 Million cell updates/sec

Title: US-09-884-465A-332

Perfect score: 4728

Sequence: 1 MQITYTDEIQVAKLAKYT.....IELRLPSGEVKKNLSDFTIA 906

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	818.5	17.3	763	10	US-09-765-272-66
2	636	13.5	796	10	US-09-765-272-66
3	431	9.1	447	10	US-09-765-272-182
4	226.5	4.8	793	9	US-09-252-088-15
5	216	4.6	715	9	US-09-252-088-16
6	197	4.2	1478	10	US-09-801-368-52
7	192	4.1	571	10	US-09-765-272-4
8	190	4.0	1596	9	US-09-902-432-4
9	181.5	3.8	1139	9	US-09-820-843A-15
10	180.5	3.8	1183	9	US-09-870-759-45
11	180.5	3.8	2368	10	US-09-815-242-5635
12	180.5	3.8	2368	10	US-09-815-242-12389
13	178	3.8	2478	10	US-09-815-242-5816
14	178	3.8	2478	10	US-09-815-242-12967
15	174	3.7	639	10	US-09-825-144-9
16	170.5	3.6	1545	9	US-09-839-996-4
17	170	3.6	1848	9	US-09-839-996-6
18	169	3.6	639	10	US-09-823-240-5
19	168.5	3.6	978	10	US-09-815-242-5456

20 168.5 3.6 1001 10 US-09-815-242-12686
21 166.5 3.5 1113 10 US-09-815-242-5836
22 166 3.5 1616 9 US-09-820-843A-16
23 166 3.5 2785 10 US-09-801-574-8
24 165.5 3.5 1181 10 US-09-870-122-23
25 164.5 3.5 1349 10 US-09-815-242-5898
26 164.5 3.5 1349 10 US-09-815-242-13137
27 164 3.5 1164 10 US-09-870-122-1
28 163.5 3.5 1881 9 US-09-998-425-3
29 163.5 3.5 1881 9 US-09-997-977-3
30 162.5 3.4 766 10 US-09-801-368-316
31 162.5 3.4 1242 9 US-09-925-299-911
32 162.5 3.4 1242 10 US-09-925-299-911
33 162.5 3.4 1781 10 US-09-738-877-3
34 162.5 3.4 26926 9 US-09-759-508B-2
35 162 3.4 1016 10 US-09-815-242-5845
36 162 3.4 2122 9 US-09-813-214A-9
37 161 3.4 1167 10 US-09-870-122-2
38 160 3.4 690 10 US-09-785-770A-4
39 160 3.4 714 10 US-09-785-770A-3
40 160 3.4 1111 10 US-09-815-242-12955
41 159 3.4 2437 10 US-09-815-242-5834
42 159 3.4 6281 10 US-09-815-242-12996
43 156.5 3.3 1125 9 US-09-974-298-114
44 156 3.3 1332 10 US-09-982-091A-4
45 156 3.3 1597 9 US-09-832-292-35

ALIGNMENTS

RESULT 1

US-09-765-272-66
; Sequence 66, Application US/09765272
; Patent No. US20020061545A1

GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272

FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 66:

SEQUENCE CHARACTERISTICS:

LENGTH: 763 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 66:

us-09-884-465a-332.rapb

21:10 2003

66

.tch 17.3%; Score 818.5; DB 10; Length 763;
 .ocal Similarity 66.1%; Pred. No. 3.2e-42;
 .nes 168; Conservative 16; Mismatches 29; Indels 41; Gaps 5;
 QY 2 QITTDDEIOVAKLAGYTTEDGYIFD-----TSWKKDLSSEAR 42
 DB 498 QITTDDEIOVAKLAGYTTEDGYIFDPRDITSDGDAYVTPHMTSHWIKDLSSEAR 557
 QY 43 AAQAAYAREKGLTPPSTHDQSGNTEARGAIAIYNRVKAAYKVPDLRMPYNLOYTVEVKN 102
 DB 558 AAQAAYAREKGLTPPSTHDQSGNTEARGAIAIYNRVKAAYKVPDLRMPYNLOYTVEVKN 617
 QY 103 GSLIIPSDHYHNIKFEWFDGLYAPKGYSLDILLATVYKYYE-----PR-----NAS 151
 DB 618 GSLIIPSDHYHNIKFEWFDGLYAPKGYSLDILLATVYKYYEHPNRPDSHDNGFGNAS 677
 QY 152 DHVRKNK-----ADQSKPDEKHEVSEPTHPSEDEKENHAGLNPSADNLYKPSDTDE 206
 DB 678 DHVRKNKQADNTNOKPSEKQPTKPEETPREKQSEKPEP-----KTEPEE 731
 QY 207 ETEEAEDTDEAE 220
 DB 732 ESPESEEPQVETE 745

RESULT 2
 US-09-765-272-56
 ; Sequence 56, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 56:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 796 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
 ;
 ; US-09-765-272-56
 Query Match 13.5%; Score 636; DB 10; Length 796;
 Best Local Similarity 45.3%; Pred. No. 4.6e-31;

Matches 140; Conservative 40; Mismatches 75; Indels 54; Gaps 8;
 QY 2 QITTDDEIOVAKLAGYTTEDGYIFD-----TSWKKDLSSEAR 42
 DB 507 QIETDEVRIAQLADRYTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIKDLSDKER 566
 QY 43 AAQAAYAREKGLTPPSTHDQSGNTEARGAIAIYNRVKAAYKVPDLRMPYNLOYTVEVKN 102
 DB 567 VAAQAYTREKGLTPSPDADYKANPTGDSAAAIYNRVKEKRIPLVRLPYHVEHTVEVKN 626
 QY 103 GSLIIPSDHYHNIKFEWFDGLYAPKGYSLDILLATVYKYYE-----PR-----NAS 151
 DB 627 GNLIPKDHYNHIAKFAWFDHHTYKAPNGYITLEDLPATIKYVHEHPDHPHSDNGWGNAS 686
 QY 152 DHVRKNKADQSKPDEKHEVSEPTHPSEDEKENHAGLNPSADNLYKPSDTDETEEE 211
 DB 687 EHV-LGKKDHSEDPNKNFADE--ERVEETPAEPE-----VPQVETEKVDAQ 730
 QY 212 AEDTDEAEI-----PCTPSIRONAMETLGLKSSLLGLTKDNNTISAEVDSLLALLKESQ 267
 DB 731 LK-----EAEVLLAKVTDSSLKANATETLAGLRNLTQIMDNNSINAEAEKULLALLKGSN 786
 QY 268 PAPIQGPQI 276
 DB 787 PSSVSKEKI 795
 RESULT 3
 US-09-765-272-182
 ; Sequence 182, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 182:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 447 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 182:
 ;
 ; US-09-765-272-182
 Query Match 9.1%; Score 431; DB 10; Length 447;
 Best Local Similarity 32.2%; Pred. No. 7e-19;

Matches 123; Conservative 40; Mismatches 91; Indels 128; Gaps 12;

QY 7 DDEIOVAKLAGITTEDGYIF-----DTS-----WIKKDSLEAERAAQA 47
 DB 144 NSNAVARSQREYTTNDGVNPNADIIETGNAYIVPRGHYHYPKSDLSASELAAGA 203
 QY 48 YAKEGLTP-----PSTHQDSGTEAKGA-----EAINRVKAAKVPPL 87
 DB 204 HLAGNMOPSQLSYSTASNDNTQVARGSTSKPANKSENLSLKLKELYDPSAQRYSSES 263
 QY 88 DRWPNLYQTVTVKNGSLIPSDYHYNKIFEFDEGLYEAPKGYSLLEDLLATVKKYVPEP 147
 DB 264 DGLVDPDAKIIIRTVNGVAIPGHDYHYPIS-----KLSALEKIAM--VPI 310
 QY 148 RNASDHVRNKADQSPDEKHEDEVSEPTHPESDEKENHAGLNPSADNLYKPSDTDEE 207
 DB 311 SGTGSTVSTN-----AKPNEVV-----SSLSGLSSN-----PSSLTTS 343
 QY 208 TEEAEEDTDEAEIPTSPIRQNAMETLTGLKSSLLGTKNNTTISAEDVSLALLKESQ 267
 DB 344 KE-----LSSASDGYIFNPKD-----IYEETA 365
 QY 268 PAPI-----QGPOIGOPTLPNNSLATPSPSLPINPGTSHKHEEDGYGFDANRI 316
 DB 366 TAVIVRHGDHYPHYPKSNQIGOPTLPNNSLATPSPSLPINPGTSHKHEEDGYGFDANRI 425
 QY 317 IAEDESGFVMSHGDSNHYFFKK 338
 DB 426 IAEDESGFVMSHGDNHNYFFKK 447

RESULT 4
 US-09-252-088-15
 ; Sequence 15, Application US/09252088
 ; Publication No. US20030031682A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRODEUR, Bernard R.
 ; APPLICANT: RIOUX, Clement
 ; APPLICANT: BOYER, Martine
 ; APPLICANT: CHARLEBOIS, Isabelle
 ; APPLICANT: HAMEL, Jose
 ; APPLICANT: MARTIN, Denis
 ; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
 ; FILE REFERENCE: 8331-9002
 ; CURRENT APPLICATION NUMBER: US/09/252,088
 ; EARLIER FILING DATE: 1999-02-18
 ; EARLIER FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 44
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15
 ; LENGTH: 793
 ; TYPE: PRT
 ; ORGANISM: group B streptococcus
 US-09-252-088-15

Query Match 4.8%; Score 226.5; DB 9; Length 793;
 Best Local Similarity 19.0%; Pred. No. 4.5e-06;
 Matches 181; Conservative 148; Mismatches 316; Indels 307; Gaps 51;

QY 65 GNTAKGAEALYNRVKAAKVPIDRMPYINQYTVTEVNGSLIIPSYDHYH-----114
 DB 18 GKAKPTNKTMDQISAEIGISAE-----QIVVKITDQGVTSHGHDYHYFNGKVPYDA 71
 QY 115 -----NKFWEDEGLYAPKGYSLLEDLLATVKKYV-----EPNASHDYKKNK--A 159
 DB 72 IISELLATDPNRYFKQSDV--INEILDGYIK-----VNGNTYVILKPGSKRKNINTRCOIA 127
 QY 160 DQSKPDDKEDHDEVSEPTHPESDEKENHAGLNPSADNLYKPSDTDETEEAEEDTDEA 219
 DB 128 EQVANGYKEAKGLAQVAHL-----SKEVAAYN-----EAKRQGYTTDDG 170
 QY 220 EIPGTPSIRQNAMETLTGLKSSLL-----LGTKNNTTISAEDVSLALLKESQ----267

DB 171 YIISPTDI-----IDDLGDALVPRGNHYHYPKDKLS--PSELAQAAYWSQKQGRG 221
 QY 268 -----PAPIQG-----POIGOPTLPNNSLATPSPSLPINPGTSHKHEH-----304
 DB 222 ARPSDYRPTPAPRRKAPIDVTPNPGQGHOPDNGGYHPAPPNPND--ASQNKHQRDEPK 279
 QY 305 -----EEDGYGFDANRIIAEDSGFVMSHGDSNHYFFKKDOLTEE 343
 DB 280 GKTFKELLQHLRLDLKYRVEEDGLIFETQVTKSNAGFYVVPVGHGHHYHIIIPRSQSLP 339
 QY 344 QIKAAQHLEEVKTSNGLSDLSHEDQYFGNAKEMKDKLKEEKIAG--IMQYGV---399
 DB 340 EMELADRY-----LAGQTEDNDSGSHSPKSDKEVTHFTLGHRIKAYKGLD 386
 QY 400 -----KRESIVVKE-----KNAIYPSGDHHDAPIDEHKPVGIGHSHSNYELFPKEE 448
 DB 387 GKPYDTDAYVFSKESHSVDKSGVTAKHGDHPHY-----IGFGEI--EQYELDEVAN 437
 QY 449 GVAKKEGNKVTGSELTNVNLLKNSTFNQNTFLANGOKRVSFSPPELEKLGKLNMLV 508
 DB 438 WV-TAKGQ-----ADELAALD-----QEQKEKPLFD-----TKKVS-----469
 QY 509 KLITPDGKV--LEKVSQKVFGEVGNITANFELDPQYLPQTF--KYTTASKDYPEVSDGT 565
 DB 470 RKVTGKQGVGYMPKQGDY-----FYARDQLDLTQIAFAEQELMKDKKHRYRD--519
 QY 566 FTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDLVRVEDEFGHGNAYLENNYK 625
 DB 520 -IVDTGIEPLRAVDVSSLPHMAGNATYDTGSSFVIPH-----IDHIIH-----560
 QY 626 VGEIKLPIPKLNQGTTRTAGNKPIPVTFMANAYLDNQST--YIVEVP-----ILKENQTD 678
 DB 561 -----VVPYSMLTR--DQIATVKYVMQHPEVRPDVMSKPGHEE 596
 QY 679 KPSILPQFKRNKAQNSKLDKEVBEPK-----TSEKVEKEKLSGTGNSSTSLLEEPTV 733
 DB 597 SGSVIPNV-----TFLDKRAGPNWQIISAEVOK--ALASGRFATPGYI-----P 642
 QY 734 DPVQKVAKFAESYGMKLENLFNMDGTIELYLPSEGEVYK--NMADTGEAPQNGENKP 792
 DB 643 DP-RDVLAK--ETFWK-----DGSFSIPRADGSSLRITKNSDLS-----QAEWQQAQ 687
 QY 793 SENGKVTG--TVENOPTENPADSLPAPNEKP---VKPENSTDNGMLNPEGVNGSDPM 847
 DB 688 ELLAKNKTGDAITDKPEKQQAQKSNZ--NQOPSEASKEEKESDDFI-----DSL 736
 QY 848 LPALEEAAPVDPVQEKLEKFTASYGLGLDS--VIFNMDGTIELRLPSEGEVI 897
 DB 737 PDYGLDRATLEDHINQLAQK-----ANIDPKYLIFQPEG--VQFYKNGELV 781

RESULT 5
 US-09-252-088-16
 ; Sequence 16, Application US/09252088
 ; Publication No. US20030031682A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRODEUR, Bernard R.
 ; APPLICANT: RIOUX, Clement
 ; APPLICANT: BOYER, Martine
 ; APPLICANT: CHARLEBOIS, Isabelle
 ; APPLICANT: HAMEL, Jose
 ; APPLICANT: MARTIN, Denis
 ; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
 ; FILE REFERENCE: 8331-9002
 ; CURRENT APPLICATION NUMBER: US/09/252,088
 ; EARLIER FILING DATE: 1999-02-18
 ; EARLIER FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 44
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 16
 ; LENGTH: 715

us-09-884-465a-332.rapb

21:10 2003

SM: group B streptococcus
J-088-16

dry Match 4.6%; Score 216; DB 9; Length 715;
Best Local Similarity 17.7%; Pred. No. 1.7e-05;
Matches 171; Conservative 127; Mismatches 271; Indels 396; Gaps 46;

QY 2 QITVDBE-----IQVAKLAKYITDCYITDTSWIKKDSLSERAAQAAYAKEGLTTP 57
DB 66 QVHLSKEEVAANRQGRYITDDGIIFS-----P 97
QY 58 STDHDSNTEAKGAEAYINRKAAYVPLDRMPYNLOYTVEVKNGLSLIIPSYDHYHNK 117
DB 98 TDIIDLDG-----AYLPHGNHYH--- 117
QY 118 FEWDEGLYEAPKGYSLDILLATVYVYEPNASHVKNKADQSKDPDEKHEVSEP 177
DB 118 -----YIPKDLSPSELAAQAYW-----SQQGRGAREPSD----- 148
QY 178 THPESDEKENHAGLPSADNLKYPSTDETEEEAEDTDEAIPGTPSIRQNAWEITG 237
DB 149 -----YRPT-----PAPG----- 156
QY 238 LKSSLLGTKNNTISAEDVSLALLESQAPIQGPIQGTPLNNSLATPSPSLPIMP 297
DB 157 -----RRKAPIDVTPNPQGHQPDNGGYYHAPPAPND- 189
QY 298 GTSEKH-----EEDGYGFANRIIAEDSGFVNSHGD 330
DB 190 -ASQKHQDEFKTKFELLQHLRLDLKYRVEEDGLIFEQTQVKNAGFYVPHGD 248
QY 331 SNHYFFKDLTEQIKAAQHLEEVKTSNGLSLSHSEHDYPCGNAKEMKOLDKLEKI 390
DB 249 HYHIPSQSLPELADRY-----LAGQEDNDGSGSEHSPSKDEYTHF 295
QY 391 AG-IMQYGV-----KRESIVNKE-----KNAIYPSGDHHDPIDEHKPVIG 435
DB 296 LGHRKAYKGLDKPYDTSDAYFSEHSISVDKSGYAKHGDHFY-----IGFG 347
QY 436 HSHSYELFKPEGVAKKGNVTGELTNVNLKNSTNNQNTLQNGKQVSYFPP 495
DB 348 EL-EQYELDEVANWV-KAQO-----ADELAAALD-----QEQKKEPLFD-- 386
QY 496 PELEKKGIMLVKLITPDGVK--LEKVGKVGEGVGNIANFELDQPLPGQTF-KYTI 552
DB 387 ---TKVS-----RKVTGDKVGYMPPKDGKY-----FYARDQLDLQIAFAEQEL 430
QY 553 ASKDYPEVSDTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFVAPVPGTDALVRVD 612
DB 431 MLKDKKRYD-----IVDTGIEPLAVDVSLSLPMHAGNATYDTSSEFVIPH-----ID 479
QY 613 EPHGNAYLENNYKVGIEKLPPELNOGTTRTAGNKIPVTFMANAYLDNOST--YIVVVP- 669
DB 480 HIB-----VVPYSWLTR-----DQIATVKYVMOHPE 505
QY 670 -----ILEKENQTPKPSILPOFKKKAQENSKLDEKVEEP-----TSEKVEKEKSETGN 720
DB 506 VRPDVWSKPGHESGSVIPNV-----TPLDKRAGMPWQIHSAREVQK-ALAEGRF 556
QY 721 STSNSTLEEVTPDQVEKAKFAESGMKLENVFNMDGTIELYPSGVEIKK-NWADF 779
DB 557 ATPDGYI-----FDP-RDVLAK--ETFWK-----DGSFISPRADGSSLRTINKSDL 600
QY 780 TGEAPOGNGENKPSNGKSVTG--TVNQPTENKPAISLEAPNEKP---VKPENSTDNQ 834
DB 601 S-----QAEWQQAQELLAKKNTGDATDTDKPEKQOQAKSNE--NOQPSSEAKSEKESDDF 654
QY 835 MLNPGNVSDDMLPALPEAPVDPVQEKLEKFTASTYGLGLDS--VIFNMDGTIELRLP 892
DB 655 I-----DSLDPYGLDRATLEDHINOLAQK-----ANIDPKYLLFQPEG-VQFYNK 698
QY 893 SCEVI 897

Db 699 NGELV 703

RESULT 6

US-09-801-368-52

Sequence 52, Application US/09801368

Patent No. US20020128250A1

GENERAL INFORMATION:

APPLICANT: Busby, Robert

APPLICANT: Cali, Brian

APPLICANT: Hecht, Peter

APPLICANT: Holtzman, Doug

APPLICANT: Madden, Kevin

APPLICANT: Maxon, Mary

APPLICANT: Milne, Todd

APPLICANT: Royer, John

APPLICANT: Salama, Sofie

APPLICANT: Sherman, Amir

APPLICANT: Silva, Jeff

APPLICANT: Summers, Eric

TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi

FILE REFERENCE: 109272.147

CURRENT APPLICATION NUMBER: US/09/801,368

CURRENT FILING DATE: 2001-03-07

PRIOR APPLICATION NUMBER: US 09/487,558

PRIOR FILING DATE: 2000-01-19

PRIOR APPLICATION NUMBER: US 60/160,587

PRIOR FILING DATE: 1999-10-20

NUMBER OF SEQ ID NOS: 440

SOFTWARE: Patent in version 3.0

SEQ ID NO 52

LENGTH: 1478

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-09-801-368-52

Query Match 4.2%; Score 197; DB 10; Length 1478;

Best Local Similarity 17.7%; Pred. No. 0.00064;

Matches 193; Conservative 139; Mismatches 307; Indels 404; Gaps 47;

QY 34 KDSLSERAAQAAYAKEKGLPPSTD-----HODSGNTEAKGAEAYINRV-----KAA 82
DB 251 QEDISNRSSTSEASLPKTS-GPSKTDENFLHSTHQKTKSASSLYRRSFTSLGSSS 309
QY 83 KVPIDRMPYNLOYTVVKNKSLI-----TPSYD-----HYHNKKEP----- 119
DB 310 SNASSAKSPNKLISIPARPHSIIESNSTLTKSASPPSPSPSIFRRHHKSSSSSLL 369
QY 120 --WFDEGLYE-----PKGYSLDILLATVYVYEPNASHVKNKADQSKDPDEK 171
DB 370 NSLFGSGIGEEAPTKPNPQGHSL-----SSENLAQK-----SKHYETNV 410
QY 172 DEVSEPTPEDEKEN--HAGLNPSADNLKYPSTDETEEEAEDTDEAIPGTPSIQ 229
DB 411 SPLKQSSLTSDKGNLWKNFKRSQIGVSPNTVAVTSQE-----TPSLKS 458
QY 230 NA-----METL-----TGLKSSLLIGTKONNTISAEDVSLALLKESQ 267
DB 459 NSSTATLVQADVNIPSPSSPPPIPKTANRSLVISTEDTPKIS-----STTASFRETY 514
QY 268 PAPIQGPQIGOPTLNNNSLATPSPSLPINCPTSHEKH-----BEDGYGFANR-- 315
DB 515 PDCINPDK-----TVP-----VPMNNQKYSVKNFLDQKFFPLKTKTGLNDSENKYI 560
QY 316 IIAEDSGGV-----MSH-----GDSNHVF 335
DB 561 LVTKDNVSVFVPLMLKSVAKLSSEKESALFKLGINHNKNTVFMHTDFDCDICAIPDITLEF 620
QY 336 FKQDL-----TEQIKAAQKHLEEVKTSNG--LDSLSHSEHDYPCGNAKEMKDL- 383
DB 621 LKSLFLNTSGKIYINDOMKLOQKPKAPLTSNNVPLKSVKSKSNRSGTSSLIASITDD 680

```

384 -----KKIEKIAGIMKQY-----GVKRESIVVNKEK 410
Db VSIYSSDITSPDEHAGSGRRYPQTPSYYYDRVNTNPTPEELNYWNKE---VLSHEE 737
QY 411 NA--IYPSGDHHDPIDEHK-----PVGIGHSHSNYELFKPEEVA---KEGKNVY 459
Db 738 NAKMVKFTSPKLEINLPDKGSKLINPTIPITENESKSFQVLRKDEGTEIDFNHRRSPY 797
QY 460 TGEEL-----TNVNLKNSFNQNTLANGQRKVSFS-----FPPELEKKL 502
Db 798 TKPELAPREAPKPPANTSPORTLSTSK-QNKPIELVRASTKISRSKSKPLPQL---L 853
QY 503 GINMLVLKITDPGKLVKSVKVPGEVGNIANFELDQPYLPQGTFTYTIASKDYP----- 558
Db 854 SSPIEASSSPD-----SLTSSYTPAST--HVLIPQYTKGAND 889
QY 559 -----EVSYDGTFTVPTSLAYKM-----ASQTIYFPFHAGDTYLRVNPQFAVEPKG 603
Db 890 VMRLKTDQDSTSP-SLKMQRKQVNRNSVSTNSIFIS----- 929
QY 604 TDALVRVDFEPHGNAYLNNYKVEIKLPIPKLNGQTTTAGNKIPVTFMANAYLDNQST 663
Db 930 -----PSLLKRGSKRV---VSSTSAADIPEENDIT 958
QY 664 YIVEVPILEKENQDKPSILPOFKNKAQENSKLDEKVEEPTSEKVEKELSETGNSTS 723
Db 959 FADAPPMDSDSDSDSS-----SSDDIINSKKKTAPETNNENKKDEKSDNSSTHS 1009
QY 724 NSTLEVPVTVQVQKVAFAESYGMKLENVLFNMDGTIELYLPGEVYKKNMADFTGEA 783
Db 1010 DEIFYDSQDQKMKRW-----FRPPEVYQNLKF---F 1043
QY 784 PQONGENKPSGKSVGTGVENQPTENKPADSL-----PEAPNEKPVPEPNT 831
Db 1044 PRAN-LDKPITEGAS-----PTSPKSLDLSLSPKNVASSRTEPSTPS-RPVPPDSY 1094
QY 832 D-----NGMLNP 838
Db 1095 EFTQDGLNGKNKP 1107

RESULT 7
US-09-765-272-4
; Sequence 4, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-765-272-4

Query Match 4.1%; Score 192; DB 10; Length 571;
Best Local Similarity 22.2%; Pred. No. 0.00037;
Matches 156; Conservative 88; Mismatches 197; Indels 262; Gaps 40;

QY 284 NSLATPSPSLPINFPGTSHRHEEDGYGFDA-----RIIADESGFVMSHGSHNYHFK 337
Db 45 SSVATPT-----KQKQVDYNNVTFVDHPVQAIQEQTPVSTKTEVQVVE 92
QY 338 KDLTEEQIKAAQHLEVKTSNGLDLSLSHEQDYPGNAKEMKDLKKIEKIAGIMKQY 397
Db 93 KPFSTELINPR---KEEQSSDSQEQLAEH-----KNLETKKEKIS----- 131
QY 398 GVKRESIVVNKEKNAIYPSGDHHDADPIDE-----HKPVGIGHSHSNYELFKPEGVA 451
Db 132 -----PKEKTV-----NLAHQDEVLSGQLNKP-----ELLYREETME 165
QY 452 KKEGNKYVTGEELNPNVNLKNSFNQNTFLANGQRVRSFPPELEKLGINK-LVKL 510
Db 166 ---TKIDFQBEI-----QENPDLAEGTVRV-----KQEGKLGKKEIVRI 202
QY 511 IT-----PDGKLVKVSGR--VFGE-----GV-----GNIANFELDQ 540
Db 203 FSVNKEEVSRIVSTTAPSPRIVEKGTQTKQVKEQPETGVEHKDVQSGAIVEPAI-Q 261
QY 541 PLYLPGQTFKTYTASRDYPSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQ--- 597
Db 262 PQLP-----EAVVSDKGEPEVQ-----TLPEAVV-----TDRGET--EVOQESPD 300
QY 598 -FAPVPGTDALVRVDFEPHGNAYLNNYKVEIK--LPKPL-NOGTTTAGNKIPVTFM 653
Db 301 TVVSDKGEPEQVAPLPEYKGN-----IEQVPETPVETKKEGQPEKT--EEVPV--- 347
QY 654 ANAYLDNQSTYIVPEPILEKENQDKPSILPQFKRN---KAQENSKLDEKVEEPTSEK- 709
Db 348 -----KPEETPVNPNBGTGTSI--QEAENPVQPAEESTTNSKV-SPDTSKN 395
QY 710 -----VEKEKLSGTGNSSTNSLLEEVTVDPVQKVAFAESYGMKLE 752
Db 396 TGEVSSNPSDSTTSVGSNKKPEHNSKENSEKTEVEVP-VNP----- 437
QY 753 NVLFNMDGTIELYLPGEVYKKNMADFTGEAPQNGENKPSGKSV-----STGTVENOPT 809
Db 438 -----NEGTV-----GTSNQETEKVPQAEETQNSGKIANENGEVSNKPSD 481
QY 810 NKP---ADSLPEAPNEKPVKPNPS-----TDNGMLNPE-----GNVSGDPMLDPALEAPVDP 860
Db 482 SKPPVEESNQPEKNGTAKPNPSNGTTSNGOTEPSPSNGNSTEDVSTESNTSNGNEE 541
QY 861 VQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGVEIKNLSD 903
Db 542 IKQNE-----LQPKKVEEPEKTELELR-----NVSD 568

RESULT 8
US-09-902-432-4
; Sequence 4, Application US/09902432
; Patent No. US20020160002A1
; GENERAL INFORMATION:
; APPLICANT: Irwin H. Gelman
; APPLICANT: Susan G. Jaken

```

QY	146	EPN	ASDHYRKNKA	ODSDPKDEK	DEHVESEPT	HP	DESEKENHAGLNP	SADNLKPS	PD	203																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
Db	18	EPEN	IFGDL	YDKSTVEE	DPNKA	YD	-----	ADNGY	YIAFNKET	GYVYDPY	68																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
QY	206	EETEEA	EDTDEAE	ITGP	TPSIRQNA	METL	GLKSL	LLGT	KNN	TISA	E 255																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
Db	69	EYD	ISQLDEN	-----	GHP	FVDEK	QENDY	LK	---	YVGN	PDYGSYBENG	EWMSGYFE 119																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
QY	256	VD	SLLALLAKESQ	PAP	IQ	-----	GPQ	I	QOPTLP	NNSLAT	PSLPS	LINPCT	SHEKHE	DG 308																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Db	120	NDQ	WIS	-	TRESQ	PTDEN	YGF	SDLP	PEV	QK	Q	SVES	DNYG	FNDL	-	PEV	QK	Q	SVES	DNYG 175																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
QY	309	YGF	DAN	R	---	I	AEDSG	FVMS	HGSD	NNHY	FFK	DLT	EEQ	I	KAQ	KH	LEV	KT	SHGL	DL 365																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
Db	176	YGF	D	N	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P	-	T	D	E	N	Y	G	F	N	D	L	P	P	E	V	Q	K	S	---	V	V	D	Q	P	S	D	D	Y	F	A	K	O	P</

253	QY	SAEVDLSIALLKESQAPIOGQPTLPNNSLATPSPSLPI-NPQTSHEKHEEDGYGF	311
498	Db	KYLKYEYNAOEDGTPESG-----TKKENGVLVYNTKEPTELTTSISGEKVWDDQNDQ	549
312	QY	DANRIIAEDSGFVMSGDGSHVFFPKDLTDEQIKAAQKHLBEVKTSHNGLDLSLSHEOD	371
550	Db	DGKR--PEKVSNNLLANG-----EKVKT-----LQVTSNWK	580
372	QY	YFNAKEMKOLDKKIE-EKIAGIMKQGVK-----RESIVNWK-----EKNALIIYPSGDH	420
591	Db	Y-----EFKDLPKYDEGKKIEYTVTEOHVADYTDINGTTITNKYTPGETSATVTKNWD	635
421	QY	HHADPIDBHKPVGCHSHSNVELFKPEGVA-----KKEGKV	458
636	Db	NNQ--DGKRPEI-----KVELY--ODGKATKTALESNNWHTWTGLDEKAKGOQV	686
459	QY	YTGELTNVNNLLKNSFTNNQNFPLANGOKRVSFSPPELKKGLGIMNLVLKITPDGKV	517
687	Db	KYTVBELTKV--KGYTHVDNDM--GNLIVTNKYTPETTSISGEKVWDDKNDQDKR	740
518	QY	LEKYSKGVFGSG-----VGNIANFELDQPLP-----GQTFKXTIAS-----KQYPEVSYD	563
741	Db	PEKVSNNLLADGEKVKTLDTVTSETNMKYEFDLPKYDEGKKIEYTVTEDHVKDY-TTDIN	799
564	QY	GTFVTPTSLAYKMASQTIFFPFHAGDTYLVY-----NPQFAPVPGYDALVRVFE	613
800	Db	GT-----TTNKYTPGETSATVTKWDDNNNDQGRPTEIKVELYQDGKAT	845
614	QY	-----FHGNAYLENNYKVGEIKLPKPLNQ---GTTTACNKPVTTFMANAYL	658
846	Db	GKTALESNNWHTWTGLDEKAKGOQVKYTVBELTKVGYTHVDNDMGNLIVTNKYT	905
659	QY	DNQSYIYVEVPILEKENTDK-----PSILPQKRKA-----QENSKLDE	699
906	Db	PETTSISGEKVWDDKNDQDKRPEKVSNNLLANGEKVKTLDTVTSNWKYEFKDLPKYDE	965
700	QY	--KVSEPTKSEKVEKLSSETGNSNSTL---EEVPTVD-----PVQEKVAK	742
966	Db	GKKIEYTVTEDHVKDYTDINGTTITNKYTPGETSATVTKWDDNNNDQGRPTEIKVEL	1025
743	QY	FAESYGMKLENYLFNMGTIELYLPDSGEVIKKNMADFT-GEAPOGNGENKPSGKYSTG	801
1026	Db	YQDGKATGKTALESNNWHTWTGLDEKAKGOQVKYTVBELTKVNGYTHVDNDMGNL	1085
802	QY	TVENQPTENKPADSL-PEAPNEK--PVKPENS	830
1086	Db	IVTNKYTPKKPNKPIYPEKPKDKTPTTPKPDHS	1117

RESULT 11

US-09-815-242-5635

Sequence 5635, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zvakind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-08-23

PRIOR APPLICATION NUMBER: 60/207,727

41:10 2003

666 VEVPILEKENOTDKPSILPQFKNRAQENSKLDKVEPKTS-----EKVEKEKL 715
 1042 TDVNAKTNEATIAITPD-ANVPTAKQAIADVQAOETAIIDANNGATTEKAAAKQO 1100
 716 SETGNSTNSILEEYPTVDPQCEKVAKEAEGMK-LENVFNMDGTIELYLPSEVKK 774
 1101 VQTEKTADTAIDGAHTNAEVE--AAKNAEIAKIEAIOPATTTKDNKAQAIATKANERT 1158
 775 NMA---DFTGE---APOGNGENKPSK-NGKSTGTVENO-----PTENK 811
 1159 ATAQODITAEIAAANANVDNAVTOANNIEAANSQNDVQAKTGTGEASIDVPTVTK 1218
 812 PADSLPEAPNEKPVKPESTDNMG 835
 1219 KATAVTDKANNITAATD---DNGV 1239

RESULT 12
 US-09-815-242-12389
 Sequence 12389, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 12389
 LENGTH: 2368
 TYPE: PRT
 ORGANISM: Staphylococcus aureus
 US-09-815-242-12389

Query Match 3.8%; Score 180.5; DB 10; Length 2368;
 Best Local Similarity 19.5%; Pred. No. 0.012;
 Matches 192; Conservative 147; Mismatches 420; Indels 225; Gaps 45;
 1 MQITYTDEIQVAKLAKYTTEDGYIFDTSWIKKSLSEAEAAQAYAKEGLTPPSTD 60
 332 LDLOFIPDNTAVADASRTTNKDGKY-YSFIDNVGLF-----SGSHLYKNRDLAPKATN 386
 61 HQ-----DSGNTKAGAEIYNRVKAAKVPLODRMPYNLOYTEVKNK---SLIIP--- 108
 387 NKEFTINTEIGNNGNFCASLKADQFKVEVTLPGQVTVVNNSLATTFPNEGDSVLKNT 446
 109 -SYDHYHNKFEWDEGLYEPKGSYLE----DLAIVKYVE-----PRN----- 149
 447 VNYQDAN-KVTFTSQGVTTARGTHTKEVLPDPKSLSKYKVNANIDTPKNIDFNEKLT 505

Query Match 3.8%; Score 180.5; DB 10; Length 2368;
 Best Local Similarity 19.5%; Pred. No. 0.012;
 Matches 192; Conservative 147; Mismatches 420; Indels 225; Gaps 45;
 1 MQITYTDEIQVAKLAKYTTEDGYIFDTSWIKKSLSEAEAAQAYAKEGLTPPSTD 60
 332 LDLOFIPDNTAVADASRTTNKDGKY-YSFIDNVGLF-----SGSHLYKNRDLAPKATN 386
 61 HQ-----DSGNTKAGAEIYNRVKAAKVPLODRMPYNLOYTEVKNK---SLIIP--- 108
 387 NKEFTINTEIGNNGNFCASLKADQFKVEVTLPGQVTVVNNSLATTFPNEGDSVLKNT 446
 109 -SYDHYHNKFEWDEGLYEPKGSYLE----DLAIVKYVE-----PRN----- 149
 447 VNYQDAN-KVTFTSQGVTTARGTHTKEVLPDPKSLSKYKVNANIDTPKNIDFNEKLT 505
 150 ---ASDHYKRNKADQSKDEKDEHSEPHSEKDEKENHAGLNPSADNLYKPSDTTE 206
 506 YRTASDIVINN-----AQEVTITADPFSVAVEMKDALQOQ--VNSQVDNSHYTTASTA 558
 207 E---TEEAEDTDEAIEPGTPIRONAMETITGLKSSLLLTGKNTTSAEVDLSALL 263
 559 EYNKQQAADNILE--DANHETANRASQAADIGLVTKIQALIDQAAIAELD---AKA 614
 264 KESQAPIQGPQIGQPTL-----PNNSLATPSPSLPINTGTSHEKHEEDGYGPOA 313
 615 QKVTAQASQKQYQDEVAALVTKINDRNAIA-----EINKQIT-----A 656
 314 NRITAEDESQFVMSHSDSNHYFFKDLTEPOIKAAQKHLEEVTSNGLDLSLSSHQDYP 373
 657 QGVTEKNGIADVLDQDVIPTTVKPOAKQDIIQAVTTKQIKKSNASQDEKDVANDKI 716
 374 G--NAKEMKOLD-----KKIEK-IAGIMKOYGVKRESIVVNEKKNAIYPSGDHHH 422
 717 GKIEKAIDDAATTAQVNAENKTAINDINOTAPATTAKAAALEEFDVQVQADQAP 776
 423 ADPIDHKPVGIGHSHNYELFPEEGVAKKEGKV-----YTGEELTNVNL 470
 777 LNPDTNEEVAEATERIN-----AAKVSQVKAIEATTAAQDLERVKNEEISKIENI 827
 471 -----LNKSTFN-----NQNTLAN--GOKRVSFSPFPPELEKLGIN--MLVK-- 509
 828 TDSQTKMDAYNEVQKQATARKTQNAVTSNATNEEVAEADAQAAQGLHDIQVYKSK 887
 510 ---LTPDGKVLKVGSKVFGGCVGNIANFELDQPLPQTPKTYTASQDYEVSDGTFT 567
 888 QEVADTKSVLDKINAIOQAKVPAADTEVENAY-----NTRKQELONS-----NASTT 937
 568 VPTSLAY-----KMASOTIFPPHAGDTYLRVNP-----QFAPVKGTDALVR 609
 938 EEKQAAVELDTKKQEAFT---NLDAATNSDVTAKDNGIAAINOVQAAATTKSDAKAE 994
 610 VFDEPHGNAYLENNYKVGKIKLPIPKLNGT---RTAGNKI-PVTFMANAYLDNQSYI 665
 995 IAQ-----KASERKTAIEAMNDSTTEEQAAKDKVQAAVVTANADIDNAAAN- 1041

Query Match 3.8%; Score 180.5; DB 10; Length 2368;
 Best Local Similarity 19.5%; Pred. No. 0.012;
 Matches 192; Conservative 147; Mismatches 420; Indels 225; Gaps 45;
 1 MQITYTDEIQVAKLAKYTTEDGYIFDTSWIKKSLSEAEAAQAYAKEGLTPPSTD 60
 332 LDLOFIPDNTAVADASRTTNKDGKY-YSFIDNVGLF-----SGSHLYKNRDLAPKATN 386
 61 HQ-----DSGNTKAGAEIYNRVKAAKVPLODRMPYNLOYTEVKNK---SLIIP--- 108
 387 NKEFTINTEIGNNGNFCASLKADQFKVEVTLPGQVTVVNNSLATTFPNEGDSVLKNT 446
 109 -SYDHYHNKFEWDEGLYEPKGSYLE----DLAIVKYVE-----PRN----- 149
 447 VNYQDAN-KVTFTSQGVTTARGTHTKEVLPDPKSLSKYKVNANIDTPKNIDFNEKLT 505
 150 ---ASDHYKRNKADQSKDEKDEHSEPHSEKDEKENHAGLNPSADNLYKPSDTTE 206
 506 YRTASDIVINN-----AQEVTITADPFSVAVEMKDALQOQ--VNSQVDNSHYTTASTA 558
 207 E---TEEAEDTDEAIEPGTPIRONAMETITGLKSSLLLTGKNTTSAEVDLSALL 263
 559 EYNKQQAADNILE--DANHETANRASQAADIGLVTKIQALIDQAAIAELD---AKA 614
 264 KESQAPIQGPQIGQPTL-----PNNSLATPSPSLPINTGTSHEKHEEDGYGPOA 313
 615 QKVTAQASQKQYQDEVAALVTKINDRNAIA-----EINKQIT-----A 656
 314 NRITAEDESQFVMSHSDSNHYFFKDLTEPOIKAAQKHLEEVTSNGLDLSLSSHQDYP 373
 657 QGVTEKNGIADVLDQDVIPTTVKPOAKQDIIQAVTTKQIKKSNASQDEKDVANDKI 716
 374 G--NAKEMKOLD-----KKIEK-IAGIMKOYGVKRESIVVNEKKNAIYPSGDHHH 422
 717 GKIEKAIDDAATTAQVNAENKTAINDINOTAPATTAKAAALEEFDVQVQADQAP 776
 423 ADPIDHKPVGIGHSHNYELFPEEGVAKKEGKV-----YTGEELTNVNL 470
 777 LNPDTNEEVAEATERIN-----AAKVSQVKAIEATTAAQDLERVKNEEISKIENI 827
 471 -----LNKSTFN-----NQNTLAN--GOKRVSFSPFPPELEKLGIN--MLVK-- 509
 828 TDSQTKMDAYNEVQKQATARKTQNAVTSNATNEEVAEADAQAAQGLHDIQVYKSK 887
 510 ---LTPDGKVLKVGSKVFGGCVGNIANFELDQPLPQTPKTYTASQDYEVSDGTFT 567
 888 QEVADTKSVLDKINAIOQAKVPAADTEVENAY-----NTRKQELONS-----NASTT 937
 568 VPTSLAY-----KMASOTIFPPHAGDTYLRVNP-----QFAPVKGTDALVR 609
 938 EEKQAAVELDTKKQEAFT---NLDAATNSDVTAKDNGIAAINOVQAAATTKSDAKAE 994
 610 VFDEPHGNAYLENNYKVGKIKLPIPKLNGT---RTAGNKI-PVTFMANAYLDNQSYI 665
 995 IAQ-----KASERKTAIEAMNDSTTEEQAAKDKVQAAVVTANADIDNAAAN- 1041

QY 150 ---ASDHVKNKADQSPDEKDEHDEYSEPHPSDEKENHAGLNPSADNLYKPSDTDE 206
Db 506 YRTASDIVNN---AQPEVTLTADPFSVAVEMKNDALQOO--VNSQVDNSHYTTASIA 558
QY 207 E---TEEEAEDTTDEAETPGTPIQONAMETUTGLKSSLLGTCKDNNTISAEDVSLALL 263
Db 559 EYNNKQOQNDNLNE-DANHVETANRASAQADGLVTKLOALIDNOQAIAELD---AKA 614
QY 264 RESQAPIQGQIGOPTL-----PNNSLATPSPSLPNTGTSHEKHEEDYGDGA 313
Db 615 QEKVTAQSKKVTQDEVAALVTKINDKNNIA-----EINKQTT-----A 656
QY 314 NRIIADSESGFVSHGDSNHYFFKDLTEQIKAAKHLEEVKTSNGLDLSLSHQDYP 373
Db 657 QGVTEKONGIAVLQDDVTTPVTKPOAKODIIOAVTRKQIKKSNASLQDEKDVANDKI 716
QY 374 G---NAKEMKLD-----KIEEK- IAGIMKQVGRKRESIVVAKENKAIIPSGDHH 422
Db 717 GKIEYKAIDIDAATNAQVEALKTKAINDINOTAPATAKAAALEEPDEVVQAQIDQAP 776
QY 423 ADPIDHKPVGHSHSNYELFKPEGVAKKEGKV-----YTGELTNVNL 470
Db 777 LNPDTTNEEVAEAIERIN-----AAKVSQVKAIEATTTAQLDERVKNEEISKIENI 827
QY 471 -----LNSTFN-----NONFTLAN--GOKRVSFSPPELEKLGIN--MLVK-- 509
Db 828 TDSOTQMDAINEVKQAARTKONTATVSNRNEEVAEADAVERAAKQGLHDQVVKSK 887
QY 510 --LITPDGKLVKSGVKGEGVGNIANFELDQYLPQGTFKYTIASKDYPEVSYDGTFT 567
Db 888 QEVADTKSKVLDKINAIQOAKVPAADTEVENAY---NTRKQEIQNS-----NASTT 937
QY 568 VPTSLAY-----KMASQTIFFPHAGDYLYLRNP-----QFAPVKGTDALYR 609
Db 938 EERQAAYTELDTKQEAR-----NLDAANTNSDVTTAKDNGIAAINOVQAATTKSKDAKAE 994
QY 610 VFDEFHGNAYLENNYKVGKILPKLNOGTT---RTAGNKI-PVTFMANAYLDNOSTYI 665
Db 995 IAO-----KASEKTAIEAMNDSTEEQQAARDKVQAVTANADIDNAAN- 1041
QY 666 VEVPILEKENOTDKPSILPOFKRKAQENSKLDEKVEEPTKS-----EKVEKEL 715
Db 1042 TDVDNAKTNEATIAATPD--ANVKTAKQAIADKQVQAQETAI DANNGATTEKAAKQ 1100
QY 716 SETGNSTNSLTLEVPVQVQKVAFAESYGNK--LENVLNMGOTIELYLPSGEVIK 774
Db 1101 VQETKTADTAIDGAHTNAEVE--AAKNAEIAKIEAIOPATTTKDNKQAIATKANERT 1158
QY 775 NMA---DFTGE---APOGGENKPS-NGKVSSTGTVENQ-----PTENK 811
Db 1159 AIAQTQDITAEETAAANAVNDNAVTOANNIEAANSQNDVQAKTTGEASIDQVTPVNK 1218
QY 812 PAOSLPAPNEKPVKPNSTNDGM 835
Db 1219 KATAVTDKNNITAATD---DNGV 1239

RESULT 13
US-09-815-242-5816
; Sequence 5816, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5816
; LENGTH: 2478
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-5816

Query Match 3.8%; Score 178; DB 10; Length 2478;
Best Local Similarity 18.9%; Pred. No. 0.018;
Matches 202; Conservative 120; Mismatches 403; Indels 344; Gaps 46;

QY 3 TTYTDETVAKLAKYTEDGYIEDTSWIKKDSL---SEAEPA---AAQAYAKEKGLT 55
Db 1433 VTQIKDQ-AVADIQG--ITADTTIKD---VAKDELATKANEOKALIAOTADATTEKEQA 1486
QY 56 PPSDHO--DSGNTAKGAEAI-----YNRVKA-----AKVPLDRMP--- 91
Db 1487 NOQVDAQLTOGQONTENAGSIDDVATKDNAIOADPIQASDVTKNARAEELTEMQNKI 1546
QY 92 ---YNLOYTVYKNGSLIIPSYDHYHNKFEWFDGLYEAPKGYSGLELLATVKYVTEPR 148
Db 1547 TEILANNETNEEKNDIGPVRAAY-----EEGLNNINAATTTGD-VTTAK-----D 1592
QY 149 NASDHVRKNKAQODSKPDEKHEDEVS-----EPHPSDEKENHAGLNPSAD-NLYK 200
Db 1593 TAVQVYQQLHAHPVVKPKAGKELDQAADKKYQIOTPNASQOEINDAKQEVDTLQNK 1652
QY 201 PSTDTETETEEAEDTTDEAETPGTPIQONAMETUTGLKSSLLGTCKDNNTISAEDV 258
Db 1653 TNVDQSSSTNEVYDNAVKECK-----AKINAVKTFSEYKKDALAKIEDAYNAKVNEADS 1706
QY 259 LIALLESQAPIQGQIGOPTLPPNNSLATPSPSLPIN-----PSTSEKHE 305
Db 1707 NASTSSEIAEAKQKLAELKOTADQNVNOATSKDDIEVQITHDLDNINDVITPTGKESAT 1766
QY 306 EDGYGF---DANRIIAEDESFGFVSHGDSNHYFFKDLTEQIKAAKHLE-EVKTSHNG 361
Db 1767 TDYIAYADOKKNISADTNA-----TQDEKQAIKOVQDQNVQTALESING 1812
QY 362 LDSLSSEHQDYPGNAKEMKDLKKEIEKTAGIMKQVGRKRESIVVAKENKAIIPSGDHH 421
Db 1813 VDN-----GDVDDALTQGAAL---DAIQVDATVTKPKANQAEVKAED-- 1852
QY 422 HADPIDHKPVGHSHSNYELFKPEGVAKKEGKVYTGTELTNVNLLKNST----- 475
Db 1853 TRSISDQSDQ-----TAEKTEALAMIKQITDOAKOG 1885
QY 476 -----FNNQNETLANGQKRVSPFPPELEKEL-----GINMLVKL 510
Db 1886 ITDATTAEVEKAKAOGLEAFDNIQDSTTEKQAEI-----ELETALDQIEAGVYNADA 1940
QY 511 ITPD-----GKVLKESGVKFGVGVGNIANFELDQYLPQGTFKYTIASKDTP 558
Db 1941 TTEKEAFTNALEDILSKATEDISDQTTNAEIAIVKNSALEQ-----LKAQRINP 1990
QY 559 EVSYDGTFTVPTSLAYKMASQTIFFPHAGDYLYR--VNPQFAPVKGTD----- 606

21.10 2003

1433 VTQKQ-ADADIG--ITADTIKD--YAKDELATKANEQKALIAQTADATEEQA 1486
56 PPSDHO--DSGNTAKAEAI--YNRKA--AKVPLDRMP-- 91
1487 NQVDAQLTOGQNONIENASIDVNTARDNAIOAIDPQASTDVKTNARAELETEMQNKI 1546
92 ---YMLQTVVEKSGSLIPSYDHYHNKPEWFEDEGLYEPKGYSLLEDLATVKKYVEPR 148
1547 TEILNNETTNEKNDIGPVRAY--EGLNINAAATTGD-VTTAK-----D 1592
149 NASDVRNKNKADQSPDEKEDHEVS-----EPHPSEDEKENHAGLNPASD-NLYK 200
1593 TAVQVQQLHANPVKPKAGKELQAAADKKTOIBOTPNASQOEINDAKQEVDELNOAK 1652
201 PSTDEETEEAEEDTDAEIPCTSIRONAWETLTKSLSLGTDK--NNTISAEVDS 258
1653 TNVDQSSSTNEYVDNAVKEGK-----AKINAVKTSEYKKALAKIEDAYNAKVNEADNS 1706
259 LLALLKESQAPIQGPQIGOPTLPNNSLATPSPSPIN-----PGTSHEKHE 305
1707 NASTSEIEAEAKQKLAELKQATADQNVNQTAKDDIEVQHNDLNDINDYTIPTGKESAT 1766
306 EDGYGF--DANRITAEDESGFVMSHGDSNHYFFKDLTEEOIKAAQKHE-EVKTSHG 361
1767 TDLAYADOKKNNISADTNA-----TODEKQOAIKQVDQNVQTALESING 1812
362 LDSLSHEQDYPGNAKEMKDLKIEKTAGIMKQGVKRESIVVKNKERNAIYPSGDH 421
1813 VDN-----GDVDDALTQKRAI--DAIQVDATVVKPANOAEVKAED-- 1852
422 HADPIDBKPVGIGHSNHYELFKPEGVAKKEGKNVYGEELTVVNNLKNST----- 475
1853 TKESIDQDQL-----TAEKTEALAMIKQITDOAKQ 1885
476 -----FNNQNTFLANGKRVSPFPPELEKKI-----GIMLVKL 510
1886 ITDATTAEVEKAKAOGLEAFDNIQIDSTEKQKAE-----ELETALDQAEAGVNVNADA 1940
511 ITPD-----GKVLKGVSGKVGEGVGNIAFELDQPYLPQGTTFYIASKDYP 558
1941 TEEKRAFTNALEDILSKATEDISQITNABIAVTKNSALEQ-----LKAQRNP 1990
559 EVSYDGTFTVPTSLAYKMASOTIFPPHAGDTYLR--VNPQFVAPKGTDA----- 606
1991 EVK-----KNALAEI-----REVYNKQIEIKKNDADASAKEIART 2036
607 -LVRVDFEFGNA-YLENNYKVGELK--LPI-----PKLNQGTTRTAGNKIPVTF 652
2027 DLGRYDFRPAKDKLTQTNAEVAELQNTIPALEAIVPQNDPDANDTNGIDNDATANS 2086
653 MANAYLDNOSTYIVVEPILEKENOTDKPSILPQFRN-----KA 691
2087 NANATPENTGQ-----PNVSETTANGKADASPTTPNNSDAATGETTATSATDANDKQPA 2141
692 QENSKLDEKVEEPKTSKVKY-EKEKSETGNSNSTLEEVPTVDVQEKVAKFAESYGMK 750
2142 NNNSSVDASTNSPTMDNDVTSKPEVESTNGTDPKPIETDNATPAESTTNNSTTTATN 2201
751 LENVLFNMDGTIELYLPSEVI-----KKNMA--DFTGEAPQNGENKP 792
2202 -ENA-----PTGSTATAPTASTEAASSADSKDNASVNDKQNAEYVNSAESQ 2248
793 SENGKYSTGTVENOPTENK-----PADSLPEAPNEKPKVPKENS----- 830
2249 STNDKVAQPKSENKAKAEKQSDSTNSQMVSESTETLPSADITEPNVPSNTSKDEEST 2308
831 ---TDNGMLNPEGVNSD-----PMLDPALEAPAVDPVQEKLEKFTAS 871
2309 NOTDAGQLKSETNVASNEADKSPSKADTEVSNKPSSTASSEAKEKMTST 2357

RESULT 15

1 EVK-----KNALAEI-----REVYNKQIEIKKNDADASAKEIART 2026
607 -LVRVDFEFGNA-YLENNYKVGELK--LPI-----PKLNQGTTRTAGNKIPVTF 652
2027 DLGRYDFRPAKDKLTQTNAEVAELQNTIPALEAIVPQNDPDANDTNGIDNDATANS 2086
653 MANAYLDNOSTYIVVEPILEKENOTDKPSILPQFRN-----KA 691
2087 NANATPENTGQ-----PNVSETTANGKADASPTTPNNSDAATGETTATSATDANDKQPA 2141
692 QENSKLDEKVEEPKTSKVKY-EKEKSETGNSNSTLEEVPTVDVQEKVAKFAESYGMK 750
2142 NNNSSVDASTNSPTMDNDVTSKPEVESTNGTDPKPIETDNATPAESTTNNSTTTATN 2201
751 LENVLFNMDGTIELYLPSEVI-----KKNMA--DFTGEAPQNGENKP 792
2202 -ENA-----PTGSTATAPTASTEAASSADSKDNASVNDKQNAEYVNSAESQ 2248
793 SENGKYSTGTVENOPTENK-----PADSLPEAPNEKPKVPKENS----- 830
2249 STNDKVAQPKSENKAKAEKQSDSTNSQMVSESTETLPSADITEPNVPSNTSKDEEST 2308
831 ---TDNGMLNPEGVNSD-----PMLDPALEAPAVDPVQEKLEKFTAS 871
2309 NOTDAGQLKSETNVASNEADKSPSKADTEVSNKPSSTASSEAKEKMTST 2357

RESULT 14
US-09-815-242-12967
Sequence 12967, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Cart, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011a
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12967
LENGTH: 2478
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12967

Query Match 3.8%; Score 178; DB 10; Length 2478;
Best Local Similarity 18.9%; Pred. No. 0.018;
Matches 202; Conservative 120; Mismatches 403; Indels 344; Gaps 46;
QY 3 ITYTDEIQVAKLAKYTTEDGYIFDTSWIKKDSL---SEAPRA---AAQAYAKEKGLT 55

```

US-09-825-144-9
; Sequence 9, Application US/09825144
; Patent No. US20020037286A1
; GENERAL INFORMATION:
; APPLICANT: Matthias Krause
; APPLICANT: Antonio S. Sechi
; APPLICANT: Frank B. Gertler
; APPLICANT: Jorgen Wehland
; TITLE OF INVENTION: Methods for Altering T-Cell and Macrophage Activation
; FILE REFERENCE: M0656/7065
; CURRENT APPLICATION NUMBER: US/09/825,144
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/194,215
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 639
; TYPE: PR1
; ORGANISM: Listeria monocytogenes
US-09-825-144-9

Query Match          3.78; Score 174; DB 10; Length 639;
Best Local Similarity 21.79; Pred. No. 0.0054;
Matches 148; Conservative 101; Mismatches 218; Indels 214; Gaps

QY 191 LNPADNLYKPPSTDEE-----TEEEAEVTDDE--AEI---PGTPIRQNAETLTGLKSS 241
DB 22 INP---DIIPANTDSESSLTNDWEKEKTEQSEVNTGPRYETAREVSSRDKELEKS 78
QY 242 LLGTTKDNNTISAEDVSLALLKESQAPIGQIGQTPGLPPLNSLATPS----- 290
DB 79 ---NKVRNTNKAD--LIAMLEKAE--KGPNNN-----NNSQTEAANAINEEASGAD 124
QY 291 -----PSLIPNCTSEKHEEDGYGDANRIIAEDSGFVMSGDSNHYFFKDL 340
DB 125 RPAIQVRRHFGPLSDSAEIKRRK-----AIASSDSELSLTYPDKPTKYNNKKV 176
QY 341 TEEQIKAAKHLFEVKTSHNGLDS-----LSHSHEQD-YPGNAKEMKOLDKKEERI--- 390
DB 177 AKESVADASE--SDLDSSMQSADSSPOPLKANQOPFFKFKKIDAGKWVRDIDENP 234
QY 391 -----AGIMKQYGVKRESIVVNEKNAIYPSGDHHHADPIDE-----HKPVG 433
DB 235 EVKKAIVDKSAGLIDQLLTKKKS---EEVNASDFF-----PPTDEELRLALPETPML 284
QY 434 IGHS-----HSNTELFEPESGVAKKEKNYTTGSELTNVYNLLKNSFTNNQNFILANGQ 487
DB 285 LGFNAPATSESPSSFFPPPP-----TDEEL-----RLALPETPMLGFNAPATS 328
QY 488 KRVSFSF-PPELEKGLGNMLVKLITPDGKLVLEKYGKVGFGGVGNIAINFELDQYLPQ 546
DB 329 EPSFEFPPTTEDELEI-----TRETASSLDSSEFTRGLAS-----LRNA 369
QY 547 TFKYTIASDKP-----EVSYDGTFTVPTSLAYKMASQTIFFPFHAGDT----- 590
DB 370 INRHQSNESDFFPIPTTEELNGRGRPTSEEFSSLSNGSDFTDENSETTEEIDRLADLR 429
QY 591 -----YLRVNQDPA---VPKGTDALVRVDFEFGNAYLENKYKVEIKLPK 635
DB 430 DRGTGKHSRNAGFLPNP-FASSPVPSLPKYSKISD----RALISDTIKTKTFFKNPSQP 484
QY 636 LN-----QCTTATAGNK-----IPVTFMANAYLDNSTVIVEPILKEKNQTDK 679
DB 485 LVNFNKKTTTATVTKKPPVVTAPKLAELPATKPOETVLRNKT-----PFIEKQAEYNK 539
QY 680 PSI-----LPQFRKRAQENSKLDERKVEEPTSEKYEKELSTGNSTS-----NSTLEEVP 731
DB 540 QSINMPSLPVIOK----EATESDKKEEMAPQTEKAVEE--SESANNAKNKNSAGIEB-- 591
QY 732 TVDPQVEKVAFAESYGMKLE 752
DB 592 -----GKLIASKADEKAKEE 607

```


GenCore version 5.1.4 ps_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:55:18 ; Search time 36 Seconds
(without alignments)
3332.506 Million cell updates/sec

Title: US-09-884-465A-332
Perfect score: 4728
Sequence: 1 MQITYDDEIQVAKLAGKYYT.....IELRLPSGEVKNLSDFIA 906

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 619202 seqs, 132417472 residues

Total number of hits satisfying chosen parameters: 619202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA-New*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pap.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pap.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pap.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pap.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pap.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pap.*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1302.5	27.5	838	5	US-09-468-656A-4
2	937.5	19.8	819	5	US-09-468-656A-10
3	818.5	17.3	763	5	US-09-765-272A-66
4	637	13.5	919	5	US-09-468-656A-8
5	636	13.5	796	5	US-09-765-272A-56
6	470	9.9	484	5	US-09-468-656A-6
7	431	9.1	447	5	US-09-765-272A-182
8	331.5	7.0	485	5	US-09-769-736-72
9	223.5	4.7	822	5	US-09-769-736-18
10	223.5	4.7	6641	6	US-10-282-122A-70580
11	223.5	4.7	10382	6	US-10-092-411A-3159
12	201	4.3	1790	6	US-10-369-493-1586
13	199.5	4.2	1098	5	US-09-797-385-8
14	197	4.2	1478	6	US-10-369-493-22154
15	195.5	4.1	1128	5	US-09-797-385-6
16	194.5	4.1	1164	5	US-09-797-385-2
17	193	4.1	1245	6	US-10-172-502-19
18	192	4.1	571	5	US-09-765-272A-4
19	192	4.1	1164	5	US-09-797-385-10
20	192	4.1	2397	6	US-10-282-122A-71232
21	191.5	4.1	3692	6	US-10-282-122A-71235
22	191.5	4.1	3696	6	US-10-092-411A-5080
23	190	4.0	4688	6	US-10-282-122A-76865
24	189.5	4.0	5005	6	US-10-282-122A-76871
25	185.5	3.9	1104	5	US-09-797-385-4
26	185.5	3.9	2109	6	US-10-369-493-6346

ALIGNMENTS

RESULT 1

US-09-468-656A-4
; Sequence 4, Application US/09468656A

; GENERAL INFORMATION:

; APPLICANT: Johnson, Leslie S.

; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus

; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural

; TITLE OF INVENTION: Motifs

; FILE REFERENCE: 469201-444

; CURRENT APPLICATION NUMBER: US/09/468,656A

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: 60/113,048

; PRIOR FILING DATE: 1998-12-21

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 4

; LENGTH: 838

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-09-468-656A-4

Query Match 27.5%; Score 1302.5; DB 5; Length 838;

Best Local Similarity 81.6%; Pred. No. 7.9e-87;

Matches 262; Conservative 2; Mismatches 6; Indels 51; Gaps 4;

QY	2	QITYDDEIQVAKLAGKYYT	DEGYIFD	-----TSWIKKDSLEAE	42
DB	518	QITYDDEIQVAKLAGKYYT	DEGYIFD	PRDITSDGDAVYTPHMTSHWIKKDSLEAE	577
QY	43	AAQAYAKEGLTPSTHDS	SGNTEAKGAEAIYRVAAKVP	LDMPYLNQYVYVKN	102
DB	578	AAQAYAKEGLTPSTHDS	SGNTEAKGAEAIYRVAAKVP	LDMPYLNQYVYVKN	637
QY	103	GSLLIPSYDHYINIKF	WDFDEGLYEPKGYSL	EDLLATVYKYYE-----PR	151
DB	638	GSLLIPSYDHYINIKF	WDFDEGLYEPKGYSL	EDLLATVYKYYEHPNPHRPHSDNGFGNAS	697
QY	152	DHYRVKNKADQSKPDE	KEHDEVSEPTHPES	DEKENHAGLNPSADNLKYPSTDTETETEE	211
DB	698	DHYRVKNKADQSKPDE	KEHDEVSEPTHPES	DEKENHAGLNPSADNLKYPSTDTETETEE	757
QY	212	AEDTTDEAEIPGT	-----PSIRONAMETLTGLK	SSLLGKTKNN	250
DB	758	AEDTTDEAEIPGT	-----PSIRONAMETLTGLK	SSLLGKTKNN	817
QY	251	TISAEVDSLLALKES	QAPAPI	271	
DB	818	TISAEVDSLLALKES	QAPAPI	838	

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272A
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Lin J. Hymel
REGISTRATION NUMBER: 45,414
REFERENCE/DOCKET NUMBER: PB340P2C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 610-5790
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-272A-66
Query Match 17.3%; Score 818.5; DB 5; Length 763;
Best Local Similarity 66.1%; Pred. No. 2.le-51;
Matches 168; Conservative 16; Mismatches 29; Indels 41; Gaps
QY 2 QITYTDEIQVAKLAGKTYTDEGYFD-----TSWIKKDSLSEAE 42
Db 498 QITYTDEIQVAKLAGKTYTDEGYFDPRDITSDGDAYTPTMTHSHWIKKDSLSEAE 557
QY 43 AAQAYAKEKGLTPPSTDRQDSGNTKAGAEAIYNRVKAARKVPLDRMPYNQYTYEVKN 102
Db 558 AAQAYAKEKGLTPPSTDRQDSGNTKAGAEAIYNRVKAARKVPLDRMPYNQYTYEVKN 617
QY 103 GSIIIPSYDHYHNKFEWDEGLYKPAKGYSLDGLATVKYVE-----NAS 151
Db 618 GSIIIPSYDHYHNKFEWDEGLYKPAKGYSLDGLATVKYVEHPNRPDSNGFGNAS 677
QY 152 DLYRVKN-----ADQSKPDEKDEHDESEPHPSDEKENHAGLNPSAONLYKPSDTDE 206
Db 678 DLYRVKNQADTNQTEKPSSEKPKTEKPEETPREKPSKESP-----KPTPEE 731
QY 207 ETEEAEDTTDEAE 220
Db 732 ESPESEEPQVETE 745
RESULT 4
US-09-468-656A-8
; Sequence 8, Application US/09468656A
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-8
Query Match 13.5%; Score 637; DB 5; Length 819;
Best Local Similarity 45.0%; Pred. No. 4.7e-38;

```

Matches 140; Conservative 41; Mismatches 76; Indels 54; Gaps 8;

QY 2 QITTTDEIOVAKLAGKTYTDEGYIFD-----TSWIKKDSLSAEAR 42
 Db 527 QIETTEDEVRIAQADKTYTSDGYIFDEHDIISDEGDAYVTPHGHSHWIGKDSLSDEK 586
 QY 43 AQAQAYAKEGLTPSPDQSDGNTAEKAGAEAIYNRVKAARKVPLDRMPYNLQYTVVEVKN 102
 Db 587 VAAQAYTKEGILPSPDADVKANPTGDSAAAIYNRVKEKRIPVLRPLVYVVEHTEVKN 646
 QY 103 GSLIIPSDYHNIKFEWFDGLYEAPKGYSLDGLLATVVKYVE-----PR-----NAS 151
 Db 647 GNLIIPIKDHYNIKFAWFDDBHYKAPNGYITLEDLFATIKYIVHPDERPHSDNGWGNAS 706
 QY 152 DHVRKNKADQSDKDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTETEER 211
 Db 707 EHV-LGKNDHSEDPKNFKADE--EPVEETPAEPE-----VPQVETEKVEAQ 750
 QY 212 AEDTTDEAEI-----PGTSPSIRQNAEMETLGLKSSLLGLTKDNNTISAEVDSLALLKESQ 267
 Db 751 LK-----EAEVLLARKVTDSSLKANATETLAGLRNNTLQIMDNNSIMAEKLLALLKGSN 806
 QY 268 PAPIOGPOIGQ 278
 Db 807 PSSVSKEKINK 817

RESULT 5
 US-09-765-272A-56
 ; Sequence 56, Application US/09765272A
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 454
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: Dell Latitude C610
 ; OPERATING SYSTEM: Windows 2000
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272A
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: OCT-30-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lin J. Hymel
 ; REGISTRATION NUMBER: 45,414
 ; REFERENCE/DOCKET NUMBER: PB340P2C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 610-5790
 ; TELEFAX: (301) 309-8439
 ; INFORMATION FOR SEQ ID NO: 56:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 796 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 56:

Query Match 13.5%; Score 636; DB 5; Length 796;
 Best Local Similarity 45.3%; Pred. No. 5.3e-38;

Matches 140; Conservative 40; Mismatches 75; Indels 54; Gaps 8;

QY 2 QITTTDEIOVAKLAGKTYTDEGYIFD-----TSWIKKDSLSAEAR 42
 Db 507 QIETTEDEVRIAQADKTYTSDGYIFDEHDIISDEGDAYVTPHGHSHWIGKDSLSDEK 566
 QY 43 AQAQAYAKEGLTPSPDQSDGNTAEKAGAEAIYNRVKAARKVPLDRMPYNLQYTVVEVKN 102
 Db 567 VAAQAYTKEGILPSPDADVKANPTGDSAAAIYNRVKEKRIPVLRPLVYVVEHTEVKN 636
 QY 103 GSLIIPSDYHNIKFEWFDGLYEAPKGYSLDGLLATVVKYVE-----PR-----NAS 151
 Db 627 GNLIIPIKDHYNIKFAWFDDBHYKAPNGYITLEDLFATIKYIVHPDERPHSDNGWGNAS 686
 QY 152 DHVRKNKADQSDKDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTETEER 211
 Db 687 EHV-LGKNDHSEDPKNFKADE--EPVEETPAEPE-----VPQVETEKVEAQ 730
 QY 212 AEDTTDEAEI-----PGTSPSIRQNAEMETLGLKSSLLGLTKDNNTISAEVDSLALLKESQ 267
 Db 731 LK-----EAEVLLARKVTDSSLKANATETLAGLRNNTLQIMDNNSIMAEKLLALLKGSN 786
 QY 268 PAPIOGPOI 276
 Db 787 PSSVSKEKI 795

RESULT 6
 US-09-468-656A-6
 ; Sequence 6, Application US/09468656A
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 ; FILE REFERENCE: 469201-444
 ; CURRENT APPLICATION NUMBER: US/09/468,656A
 ; CURRENT FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 484
 ; TYPE: PPT
 ; ORGANISM: Streptococcus pneumoniae
 ; US-09-468-656A-6

Query Match 9.9%; Score 470; DB 5; Length 484;

Best Local Similarity 33.6%; Pred. No. 3.6e-26;

Matches 131; Conservative 40; Mismatches 91; Indels 128; Gaps 12;

QY 7 DDEIOVAKLAGKTYTDEGYIF-----DTS-----WIKKDSLSAEARAAQA 47
 Db 168 NSNVAVARSQGRYITNDGYVFNPAIDTGNAYVPHGGHYHYIPKDSLSASELAAGA 227
 QY 48 YAKEGLTP-----PSTDHQDSGNTAEKGA-----EAIYNRVKAARKVPL 87
 Db 228 HLAGKNQPSQISYSTASDNTQSVANGSTKPAKNSLQSLKELYDSPSAQRYSSES 287
 QY 88 DRMPYNLQYTVVEKNGSLIIPSDYHNIKFEWFDGLYEAPKGYSLDGLLATVVKYVEP 147
 Db 288 DGLVFPDPAKIISRTNPNGVAIPHGDHYHFIPYS-----KLSALEEKIARM---VPI 334
 QY 148 RNASDHVRKNKADQSDKDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTETE 207
 Db 335 SCTGSTVSTN-----AKPNEV-----SSLGSLSN---PSSLTTS 367
 QY 208 TEEAEADTTDEAEIPGTPSIRQNAEMETLGLKSSLLGLTKDNNTISAEVDSLALLKESQ 267
 Db 368 KE-----LSSASDGYIFNPKD-----IVEETA 389

41.11 2003

QY 208 TEEAEEDTDAEIPGTPSIRONAMETILGLKSSLLGLTKDNNTISAEDVSLALLKESQ 267
 Db 344 KE-----LSSASDGYIFNPKD-----IVEETA 365
 QY 268 PAPI-----QGPOIGQPTLPNNSLATPSPSLPINPOTSHKHEEDGYGFDANRI 316
 Db 366 TAYIVRHGDHFHYIPKSNQIGQPTLPNNSLATPSPSLPINPOTSHKHEEDGYGFDANRI 425
 QY 317 IAEDESGFVMSHGDSNHYFFKK 338
 Db 426 IAEDESGFVMSHGDSNHYFFKK 447

RESULT 8

US-09-769-736-72
 ; Sequence 72, Application US/09769736
 ; GENERAL INFORMATION:
 ; APPLICANT: Microbial Technics Limited
 ; APPLICANT: Le Page, Richard WF
 ; APPLICANT: Wells, Jeremy M
 ; APPLICANT: Hanniffy, Sean B
 ; TITLE OF INVENTION: Proteins
 ; FILE REFERENCE: PWC/P21089WC
 ; CURRENT APPLICATION NUMBER: US/09/769,736
 ; CURRENT FILING DATE: 2003-02-14
 ; PRIOR APPLICATION NUMBER: GB 9816335.5
 ; PRIOR FILING DATE: 1998-07-27
 ; PRIOR APPLICATION NUMBER: US 60/125163
 ; PRIOR FILING DATE: 1999-03-19
 ; NUMBER OF SEQ ID NOS: 212
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 72
 ; LENGTH: 485
 ; TYPE: PRT
 ; ORGANISM: Streptococcus agalactiae
 US-09-769-736-72

Query Match 7.08; Score 331.5; DB 5; Length 485;
 Best Local Similarity 23.88; Pred. No. 5.2e-16;
 Matches 95; Conservative 48; Mismatches 89; Indels 167; Gaps 9;

QY 2 QITYTDDIYVAKLAGKYTTEDGYIFDTS-----WIKKDSLSAEAR 42
 Db 94 QIVYSAQIEEAKKAGKYTSDGYIFDANKDKDTGTGVIPHMTHEHWVPKDLSESEL 153
 QY 43 AAAQAYA-----49
 Db 154 KAAQEFSLKSEANQDKPKTGAQIEAIEPKAIVKPEDLLFGIAQATYKNGTFVIP 213
 QY 50 -----KEKGL-----54
 Db 214 HKDHYHYVELKWFDEBKDLADSDKTSYLEDYLATAKYMMHPRKPVGKGDAEYK 273
 QY 55 -----TPSTDHODSGNTEAKGAZA--IYNNRKAARKVPLDRMPYNIQYTVKN 102
 Db 274 EKDSNKADEKAPDNTKSTNSGDKNLAAAEVFKOAKPEKIVPLDKIAAHMAYAVGED 333
 QY 103 GSLIIPSYDHYHNKFEWFE-GLYEPKGYSLDGLATVYVEPRNA-----S 151
 Db 334 DQILVPHDHYHNVPMWFDKGLKAKPEGYTLQQLFSTIKYMEHPNLPKKGWGHDS 393
 QY 152 DHYRKNKAODSK---PDEKHEDEVSEPHPEDEKENHAGLNPSADNLYKPSDTEET 208
 Db 394 DHNKGSKMKNKANYAPDEE-----PEDSGKVTH---NYGFYDVNGSDEEPE 439
 QY 209 EEEAEEDTDAEIPGTPSIRONAMETILGLKSSLLGLTK 247
 Db 440 KQDESELDETELMQAQNAKKGNDQSGFEKQLIQLSNK 478

RESULT 9
 US-09-769-736-18
 ; Sequence 18, Application US/09769736

QY 317 IAEDESGFVMSHGDSNHYFFKKOLTEQIK 346
 Db 450 IAEDESGFVMSHGDSNHYFFKKOLTEQIK 479

RESULT 7

US-09-765-272A-182
 ; Sequence 182, Application US/09765272A
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
 ; NUMBER OF SEQUENCES: 454
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: Dell Latitude C610
 ; OPERATING SYSTEM: Windows 2000
 ; SOFTWARE: ASCII text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272A
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: OCT-30-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lid J. Hymel
 ; REGISTRATION NUMBER: 45,414
 ; REFERENCE/DOCKET NUMBER: PB340P2C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 610-5790
 ; TELEFAX: (301) 309-8439
 ; INFORMATION FOR SEQ ID NO: 182:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 447 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 182:
 US-09-765-272A-182

Query Match 9.18; Score 431; DB 5; Length 447;
 Best Local Similarity 32.28; Pred. No. 2.3e-23;
 Matches 123; Conservative 40; Mismatches 91; Indels 128; Gaps 12;

QY 7 DDEIQVAKLAGKYTTEDGYIF-----DTS-----WIKKDSLSAEARAAQA 47
 Db 144 NSNVAARSQRYTNDGVFNPAIDIEDTGNAYIVPHGHHYIPKSDLSASELAAGA 203
 QY 48 YAKEKGLTP-----PSTDHODSGNTEAKGA-----EAYNNRKAARKVPL 87
 Db 204 HLAGKNMOPSQLSYSTASDNTOSVAKGSKPANKSENQSLKELYDSPSQORSES 263
 QY 88 DRMPYNLYTVKNGSLIIPSYDHYHNKFEWFEGLYEPKGYSLDGLATVYVEPR 147
 Db 264 DGLVDPAPKIIISRTNGVAIPGHGHHYIPYS-----KLSALEEKIARN---VPI 310
 QY 148 RNASDHYRKNKADQSKPDEKHEDEVSEPHPEDEKENHAGLNPSADNLYKPSDTEET 207
 Db 311 SGTGSTVSTN-----AKPNEV-----SSLSGLSSN---PSSLTTS 343

GENERAL INFORMATION:
 APPLICANT: Microbial Technics Limited
 APPLICANT: Le Page, Richard W
 APPLICANT: Wells, Jeremy M
 APPLICANT: Hanniffy, Sean B
 TITLE OF INVENTION: Proteins
 FILE REFERENCE: PNC/P21089wo
 CURRENT APPLICATION NUMBER: US/09/769,736
 CURRENT FILING DATE: 2003-02-14
 PRIOR APPLICATION NUMBER: GB 9816335.5
 PRIOR FILING DATE: 1998-07-27
 PRIOR APPLICATION NUMBER: US 60/125163
 PRIOR FILING DATE: 1999-03-19
 NUMBER OF SEQ ID NOS: 212
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 18
 LENGTH: 822
 TYPE: PRT
 ORGANISM: Streptococcus agalactiae
 US-09-769-736-18

Query Match 4.7%; Score 223.5; DB 5; Length 822;
 Best Local Similarity 19.0%; Pred. No. 9.8e-08;
 Matches 186; Conservative 143; Mismatches 319; Indels 301; Gaps 52;

QY 65 GNTKAGAEATYRKAARKVPLDRMPYNLQYTVKNGSLIIPSDYH-----114
 DB 47 GKVKAPKTKNTWDOISAEIGISAE-----QIVVKITDQYVTSBGDHYHFGYNGKVPYDA 100
 QY 115 -----NIKPFWDEGLYAPKGYSLDLATVYIV--EPRNASDHVRKMK--A 159
 DB 101 IISEELMTDPNTHPKQSDV--INEILDGVIVK---VNGNYVTVLKPGSKRKNIRTKQIA 156
 QY 160 DQSKPDEKDEHVESEPHSPDEKENAGLNPSADNLYKSTDTETEERAEEDTDEA 219
 DB 157 EQVAKTEKAEKGLAQVAHL-----SKEEVAVN-----EAKRGYITDGG 199
 QY 220 EIPGTPSIRQAMETITGLKSSLL-----LGTKNNTISAEDVSLALIKESQ----267
 DB 200 YIFSPTDI-----IDLDGDAYLVPHGNHYHYPKDKLS--PSELAAQAYWSQKRG 250
 QY 268 -----PAPIQG-----PQIGQTLPNNSLATSPSLPIMPPTSHEKH-----304
 DB 251 ARPSDYRPTAPGRRKAPIDVTPNPGQHQPDNGYHPAPPNDP--ASQNKHQRDEPK 308
 QY 305 -----EDGCGFDMRRIIAEDSGFVMSHSDSNHYFFKKDLTEE 343
 DB 309 GKTFKELLHLRLDLKYRVEEDGLIFETQVIKSNAGYVYPHGDHYHIIIPRSQSLP 368
 QY 344 QIKAAQKHLEEVKTSNGLSLSHSQDYPGNAKEMKDLDKTEEKIAG--IMKQYGV---399
 DB 369 EMELADRYLAG-QTDON--DSGSDHS-----KPSDKVTHFTFLHRIKAYKGLD 415
 QY 400 -----KRESIVNKE-----KNAIYPSGDHHDAPIDEKHPGVGSHSNTELEKPEE 448
 DB 416 GRPYDTSAYVFSKESHSYKAGKGDHFI-----IGFGL-BOYELDEVAN 466
 QY 449 GVAKEGKNYVTGEBLTNVNLLKNSTFNQNTLANGOKRVGSFPPPELEKLGIMLV 508
 DB 467 WV-KAQQ-----ADELVAUD-----QEGKKEKPLFD-----TKKVS-----498
 QY 509 KLITPDGVK--LEKSVKGVGEGVGNIANFELDQPLPGQTFKTYTASQDPEVSYDGT 566
 DB 499 RYVTKGKGVGIMPDKGKY-----PYRYQLDLTQI--AFAEQELMLKDKHYRYD---548
 QY 567 TVPSTSLAYKASQTIPTFFIAGDTYLRVNPQPAVPGKTDALRVFDEFGHNAV--LENNY 624
 DB 549 IVDTGIEPLAVDSLSLPMAGNATYDTGSSFVIPH-----IDHIVVYVPSWLTNRN-599
 QY 625 KVGSEIKLPIPLKNGTTRTAGNKIPVTFMANAVLDNGSTVIVEVP-----ILEKENQTDK 679
 DB 600 QIATIK-----YVMQHPVRRPDVNSKPGHEES 626

QY 680 PSILPQFKRKAQENSKLDEKVEPK-----TSEKVEKEKLSSTNSTLEEVPTVD 734
 DB 627 GSVIPNV-----TPDKRAGHPNQIHSAAEVOK-ALAEGRFAAPDGYI-----ED 672
 QY 735 PVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYIKK-NMADFTGEAPQNGENKPS 793
 DB 673 P-RDVLAK--ETFWK-----DGSFSIPRADGSSLRITINKSDLSQAEWQQAQELLAK 721
 QY 794 ENKGVSTGTVENQPTENKPADSLPEAPNEKP---VKPENSTDNGLNPEGNVSGDPMLDP 850
 DB 722 KNAGDATDT--DKPEEQQADKSNE--NOQSEASKESKESDDFI-----DSLDPY 768
 QY 851 ALEAPAVDPVQEKLEKFTASYGLGLDS--VIFNMDGTIELRLPSGEVI 897
 DB 769 GLDRATLEDHINQLAQK-----ANIDPKYLIFQPEG-VQFYNNKNGELV 810

RESULT 10
 US-10-282-122A-70580
 Sequence 70580, Application US/10282122A
 GENERAL INFORMATION:
 APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.
 APPLICANT: Xu, H.
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: ELITRA 034A
 CURRENT APPLICATION NUMBER: US/10/282,122A
 CURRENT FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/230,335
 PRIOR FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: 60/230,347
 PRIOR FILING DATE: 2000-09-09
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 Remaining Prior Application data removed - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 70580
 LENGTH: 6641
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 FEATURE:
 NAME/KEY: MISC-FEATURE
 LOCATION: (6636)..(6636)
 OTHER INFORMATION: X-any amino acid
 US-10-282-122A-70580

Query Match 4.7%; Score 223.5; DB 6; Length 6641;
 Best Local Similarity 19.0%; Pred. No. 2.6e-06;
 Matches 201; Conservative 169; Mismatches 409; Indels 279; Gaps 47;

11:11 2003

RESULT 11
US-10-092-411a-3159
Sequence 3159, Application US/10092411a
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-101
CURRENT APPLICATION NUMBER: US/10/092,411a
CURRENT FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: US 09/134,001
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5676
SEQ ID NO 3159
LENGTH: 10182
TYPE: PAT
ORGANISM: Staphylococcus epidermidis
US-10-092-411a-3159

Query Match 4.7%; Score 223.5; DB 6; Length 10182;
Best Local Similarity 19.0%; Pred. No. 5.1e-06;
Matches 201; Conservative 169; Mismatches 409; Indels 279; Gaps 47;

QY 37 LSEARAAQAYAKEGLTPPSTDDHSGNTEAKGAEAIYNRVKAAYKVPDMPYINQ- 95
1704 LTRAKEDAVASINNLGTLNQPKENAVGAQTRDOVANKLRDAE--ALDQSMOTLRD 1761
QY 96 -----YTVKNGSLIPSYDHYHNKFEWDEGL----- 125
1762 LVNNQNAIHSTNYFNEDSTQKNTYDAIDNGSTYITG-QHNPELNKSTIDQISRINTA 1820
QY 126 -----YAPGKYSLEDLATVKKYVPRNASDVRKNKADQDSKPDE----- 167
1821 KNDLHGVEKLQDKGTANOE-IGOLGYLNDPQKSGEESLVNGSNTSRVEEHLNEAKSLN 1879
QY 168 ---DKEHDEVESEPHPE-----SDEKENHAGLNP-----ADNLY-----KPSDTDEETEE 211
1880 NAMQLRDKVAKETNVKSSDIYNDSTEHQGYDQALQAEANIINEIGNPTLNKSEIQK 1939
QY 212 AEDTDEAEIPGTPSIRQNAETTLGLKSSLLGTDKNTTISAEDVSLALIKESQAPAI 271
1940 LQQLTD-----AQNA-----LOGSHLLEAKNNAITG-INKLALNDAQORQAI 1982
QY 272 QGPQIGQPTLP-----NNSLATPSPSLPINPGTSHEKHEEDGY-----G 310
1983 ENVO-AQOTIPAVNQOGLTLDREINTAMQALRKVQGNVHOOSNYFNEDEQPKHNYDNS 2041
QY 311 FDNRIIAEDESFGVMHSGDSNHYFFKDLTEQIKAAOK-HLEEVKTSH--NGLDLSL- 366
2042 VQAGOTIIDKLQDPIMNKNEIEQAINQINTOTALSGENKLTQDSTNRQIEGLSSLT 2101
QY 367 ---SHEQDYPGNAKEMKDLKKI-----EKKIA-----GIMQYGVKRESIVVYKKNKAI 413
2102 AQINAEKDLVQAKTRDVAQKAAKEINSAMSLRDGQONKEDIKRSAYINADPTKV 2161
QY 414 IYPSGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKGNKVTYTGELTNVYLLKN 473
2162 TAYDQALQNAENIINATP-----NVELNK-----ATIEQALSRVQAOQDLDGVQOOLAN 2210
QY 474 S-----TFNNQNTFLANGQKRVSFSPPELEKLGINMLVKLTIPDGKLVKYSV 523
2211 AKQOATQVINGLN-SLNDGQKR-----ELNLLINSANTRTKVOEELNKATEL 2256
QY 524 ---KVFGEVGNITANFELDQPLPGOTFKYTIASKDYPEV-SYDGTFTVPTSLAYKNAS 578
2257 NHAMEALNSVQNVDOVKQSSNYV-----NEDQPEQHNYDN-----AVNEAQ 2298
QY 579 QTIFFPFFHAGDTYLRVNPQFAPVPGKTDALVRVDFEHGNAVLENNYKVGEIKL----- 631
2299 ATI-----NNAQVLDKLAERLTQVNTTKDALHGAQKLTQDQAAETGIRGLTSLN 2352
QY 632 -----PIPLNQGTRTAGNKI--PVTMANAYL-----DNQSTYIIVE----- 667
2353 EPQKNAEVAKYTAATTRDEVNRIQEAETLDTAMGLGRKSIKDKNTKSKYINEDHDQ 2412
QY 668 -----VPILEKQNTDPSILPQKRNKAQENSKL--DERVEEPKTEK---VE 711
2413 QQAYDNVAVNNAQVIDETQATLSDDINQLANAVTQAKSNLHGDTKLQHKDSAKOTIAQ 2472
QY 712 KEKLSGTGNTSSTLEVTVPVQKFAESYGMKLENVLFNMDGTIELY---LPS 768
2473 LQINSAQKEMEDSLIDNESTRITQVQHDLE-----AQALDGLMGALKESIKDITNIVSN 2527
QY 769 GEVI-----KKNADFTGEAPQG---NGENKPSN-KGVSTG--TVENOPTENKPADSLPE 818
2528 GNYINAEPSKQAYDAVQNAQNIINGNTOPTINKNGNVTATQVKNYTKDALDGDHRL 2587
QY 819 APN---EKVPKSTNDGMLNPEGVSGDPMLDPALEAPAVDPVQEKLEKFTASYIGL 876
2588 ARNNAQOTIRNLNNAQKDAEKL-----VNSASTLEQVQOOLN---TAQQ---L 2633
QY 877 DSVIFNMDGTIELR-----LPSGEVIRKNSLDFI 905
2634 DNANGELQSTIAKQOVADSKYLNEDPQIKQNYDDAV 2671

QY 37 LSEARAAQAYAKEGLTPPSTDDHSGNTEAKGAEAIYNRVKAAYKVPDMPYINQ- 95
1704 LTRAKEDAVASINNLGTLNQPKENAVGAQTRDOVANKLRDAE--ALDQSMOTLRD 1761
QY 96 -----YTVKNGSLIPSYDHYHNKFEWDEGL----- 125
1762 LVNNQNAIHSTNYFNEDSTQKNTYDAIDNGSTYITG-QHNPELNKSTIDQISRINTA 1820
QY 126 -----YAPGKYSLEDLATVKKYVPRNASDVRKNKADQDSKPDE----- 167
1821 KNDLHGVEKLQDKGTANOE-IGOLGYLNDPQKSGEESLVNGSNTSRVEEHLNEAKSLN 1879
QY 168 ---DKEHDEVESEPHPE-----SDEKENHAGLNP-----ADNLY-----KPSDTDEETEE 211
1880 NAMQLRDKVAKETNVKSSDIYNDSTEHQGYDQALQAEANIINEIGNPTLNKSEIQK 1939
QY 212 AEDTDEAEIPGTPSIRQNAETTLGLKSSLLGTDKNTTISAEDVSLALIKESQAPAI 271
1940 LQQLTD-----AQNA-----LOGSHLLEAKNNAITG-INKLALNDAQORQAI 1982
QY 272 QGPQIGQPTLP-----NNSLATPSPSLPINPGTSHEKHEEDGY-----G 310
1983 ENVO-AQOTIPAVNQOGLTLDREINTAMQALRKVQGNVHOOSNYFNEDEQPKHNYDNS 2041
QY 311 FDNRIIAEDESFGVMHSGDSNHYFFKDLTEQIKAAOK-HLEEVKTSH--NGLDLSL- 366
2042 VQAGOTIIDKLQDPIMNKNEIEQAINQINTOTALSGENKLTQDSTNRQIEGLSSLT 2101
QY 367 ---SHEQDYPGNAKEMKDLKKI-----EKKIA-----GIMQYGVKRESIVVYKKNKAI 413
2102 AQINAEKDLVQAKTRDVAQKAAKEINSAMSLRDGQONKEDIKRSAYINADPTKV 2161
QY 414 IYPSGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKGNKVTYTGELTNVYLLKN 473
2162 TAYDQALQNAENIINATP-----NVELNK-----ATIEQALSRVQAOQDLDGVQOOLAN 2210
QY 474 S-----TFNNQNTFLANGQKRVSFSPPELEKLGINMLVKLTIPDGKLVKYSV 523
2211 AKQOATQVINGLN-SLNDGQKR-----ELNLLINSANTRTKVOEELNKATEL 2256
QY 524 ---KVFGEVGNITANFELDQPLPGOTFKYTIASKDYPEV-SYDGTFTVPTSLAYKNAS 578
2257 NHAMEALNSVQNVDOVKQSSNYV-----NEDQPEQHNYDN-----AVNEAQ 2298
QY 579 QTIFFPFFHAGDTYLRVNPQFAPVPGKTDALVRVDFEHGNAVLENNYKVGEIKL----- 631
2299 ATI-----NNAQVLDKLAERLTQVNTTKDALHGAQKLTQDQAAETGIRGLTSLN 2352
QY 632 -----PIPLNQGTRTAGNKI--PVTMANAYL-----DNQSTYIIVE----- 667
2353 EPQKNAEVAKYTAATTRDEVNRIQEAETLDTAMGLGRKSIKDKNTKSKYINEDHDQ 2412
QY 668 -----VPILEKQNTDPSILPQKRNKAQENSKL--DERVEEPKTEK---VE 711
2413 QQAYDNVAVNNAQVIDETQATLSDDINQLANAVTQAKSNLHGDTKLQHKDSAKOTIAQ 2472
QY 712 KEKLSGTGNTSSTLEVTVPVQKFAESYGMKLENVLFNMDGTIELY---LPS 768
2473 LQINSAQKEMEDSLIDNESTRITQVQHDLE-----AQALDGLMGALKESIKDITNIVSN 2527
QY 769 GEVI-----KKNADFTGEAPQG---NGENKPSN-KGVSTG--TVENOPTENKPADSLPE 818
2528 GNYINAEPSKQAYDAVQNAQNIINGNTOPTINKNGNVTATQVKNYTKDALDGDHRL 2587
QY 819 APN---EKVPKSTNDGMLNPEGVSGDPMLDPALEAPAVDPVQEKLEKFTASYIGL 876
2588 ARNNAQOTIRNLNNAQKDAEKL-----VNSASTLEQVQOOLN---TAQQ---L 2633
QY 877 DSVIFNMDGTIELR-----LPSGEVIRKNSLDFI 905
2634 DNANGELQSTIAKQOVADSKYLNEDPQIKQNYDDAV 2671

Db 5867 ATI-----NNNAOVLKDLATERLTQTNTTKDALHGAQKLTQOQAAETGIRGLTSLN 5920
 QY 632 -----PIPLNGOTRTAGNKI--PVTMANAYL-----DNOSTYIYE-----667
 Db 5921 EPORNAEAVKATTAATRDVNRIRQBATLDTAMGLRKSIOKNDTKSSKYINEDHDO 5980
 QY 668 -----VPILEKNOTDKPSLPQKRNKAQENSKL--DEKVEPKYSEK---VE 711
 Db 5981 QOAYDNVNAOQVIDETQATLSSDTINOLANAVTQAKSNLHGDTKLOHDKDSAKOTIAQ 6040
 QY 712 KEKSETGNTSNTLEEVTPDPOKPAESYGMLEWLFNMDCTIELY---LPS 768
 Db 6041 LONLSAQKMSDLINESTRTQVQDHLTE-----AQALDGLMGALKESIKTYTWNISN 6095
 QY 769 GEV-----KKNMADFTGEAPOG--NGENKPSN--GKYSTG--TVENOPTENKPADSLPE 818
 Db 6096 GNYINAEPSSKQAYDAVQNAQNIINGTNOPTINKGNVTATQTVKTKDALDGDHRLSE 6155
 QY 819 APN--ERKVPKPSNDGMLNPGVSGDPMPLDPALEAPADVPDPOKLEKFTASYGLGL 876
 Db 6156 AKNNANOTIRNLSNNAQDAEKNL-----VNSASTLEQVOQNLO--TAQO---L 6201
 QY 877 DSVTFNMDGTIELR-----LPSGEVKKNLSDPI 905
 Db 6202 DNANGELRQSTAKKQVKAQSKYLNEDPOIKQNTDDAV 6239

RESULT 12

US-10-369-493-1586
 ; Sequence 1586, Application US/10369493
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 1586
 ; LENGTH: 1790
 ; TYPE: PRT
 ; ORGANISM: Saccharomyces cerevisiae
 US-10-369-493-1586

Query Match 4.38; Score 201; DB 6; Length 1790;
 Best Local Similarity 19.36; Pred. No. 1.5e-05;
 Matches 183; Conservative 132; Mismatches 328; Indels 304; Gaps 40;
 QY 67 TEAKGAEIYNRVKAAKVPDLMRYNLYQYTVKNGSLIIPSYD-----HYHNKIFE 119
 Db 650 TTKLGDNYASIKQFKK-----DSYFSKVDNEDSILAPDELDTGLPKVYFTYFIQ 702
 QY 120 WFDEGLYKAPKYSLEDLLATVYKVEPRNASDHYKKNKADQSPDDEKHEDEV-----174
 Db 703 LFENIYRI-----RTALSH-----DPDEEPINKISFEEVERKLOR 737
 QY 175 -----SEPTHPESDEK-----ENHAGLNPDSADNLYKPSDTDEE 207
 Db 738 QCTKLAGEITSQTETSTHEHNTKLTALTHEHKLDEKYLINSSHSL--KENFSILE 796
 QY 208 TE--EEAEDTTDEAIEPTGSIRQNAEMETITGLKSLLLGTDKNNTTISAFVDSLLALLKES 266
 Db 797 TELKNYRDSILDE-----MTQLRDVLETKDKENQATALLEYKSTI-----HK 836
 QY 267 QPAPIQGOPIQOPLPNNSLATPSPSLPNPCTSHKHEEDG-----YGFANDRIITA 318

Db 837 QEDSIKTEKLEGTI-----LSQKKKAEDGINKMGKDLFALSREMQAV 879
 QY 319 EDESGFVMSHGS---NHFFKDLTEB-----QITAAQHLEEVKTSNGL-----DSL 365
 Db 880 EENCKNLQKEDKSNVNHQKETSRLKEDIAAKITEIKAINLEEMKIQCNLNSKEKEHI 939
 QY 366 SSHEDQYPGNAKEMKDLKKEIEKIAGIMKOY---GVKRESIV--VNKEKN--AIYPSGD 419
 Db 940 SKELVEYKRSQSHDNLVAKLTEKLSANNYKQMAENESLIRAVEESKNESQIOLSNL 999
 QY 420 HHADPIDDEHK---PVGIGHSHSNYELFK-----PEEGVAKKEGKNYVTEELTNV 467
 Db 1000 ONKIDMSQERENFOIERGSEIKETKIEQLKTKTISLEQTKEEIISKSDSK-----DEYESQ 1055
 QY 468 VNLKN-----STFNQNFLLANGOKRYSFSPPELE--KKLGINMLVKLTTPGKVLEK 520
 Db 1056 ISLLKEKLETATANDENVNKISELTKTRELEAEALAYKNLKNLETKLETSE--KALKE 1114
 QY 521 V-----SGKYFGEVGNIANFELDQYLPFGQTPKY--TIASKDYP 558
 Db 1115 VKEENEHLKEEKIQLEKEATETKQOLNSLRANLESLEKEHEDELAQAKKYEEOIANK---1171
 QY 559 EYVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQF--AVPKGTDALVRVDFEFGN 617
 Db 1172 ERQYN-----EISQLNDEITSTOOENESIKKKDELEGE 1206
 QY 618 AVLENNYKVGELKPIPKLNQGTTRTAGNKIPVTMANAYLDNQSTYIYVEVPILEKENQT 677
 Db 1207 V-----KAMKSTSEQSN-----LKKSEIDALNLQIKE---LAKKNET 1241
 QY 678 DKPSILPOPK-----RNKAQENSKLDEKVE--EPKTSKVEKEKLESE---717
 Db 1242 NEASLLESIKSVESVTVKIKELQDECNFKEKVESELEDKLASEDKNSKYLELOKESKI 1301
 QY 718 -----TGNSTSNSTLEEVTPDPOK-----VAKFAESYGMLEWLFNM 758
 Db 1302 KEELDAKTELKIQLEKINTLAKAKESSELSRLKTSSEKKNABEOLKLN-----1356
 QY 759 DGTIELYPSGEVYKKNMADFTGEAPOGNGENKPSNGKSVSTGTVENOPTENKPADSLPE 818
 Db 1357 -----ETQIKNQA-----FEKERLLNEGS--STTOYSEKINTLEDELIR 1396
 QY 819 APNEKVPKPSNDGMLNPGVSGDPMPLDPALEAPADVPDPOKLE 865
 Db 1397 LQNELKAKEDIDTRSELEKVSLSNDELLE-----EKONTIKSLQDEI 1440

RESULT 13

US-09-797-385-8
 ; Sequence 8, Application US/09797385
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai, Joseph Y.
 ; APPLICANT: Blake, Milan S.
 ; TITLE OF INVENTION: Non-IgA Fc Binding Forms of the Group B
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC Compatible
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/797,385
 ; FILING DATE: 01-Mar-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:

us-09-884-465a-332.rapn

.21:11 2003

APPLICATION NUMBER: US 08/923,992
 FILING DATE: 05-SEP-1997
 APPLICATION NUMBER: US 60/024,707
 FILING DATE: (VII)
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REGISTRATION NUMBER: 32,893
 REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1098 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Query Match 4.28; Score 199.5; DB 5; Length 1098;
 Best Local Similarity 17.69; Pred. No. 8.8e-06;
 Matches 157; Conservative 125; Mismatches 289; Indels 321; Gaps 34;

166 DEKHEVSEPTHP---ESDEKENHAGLNPSADNLYKPSDTEETEEAEDTDEAEP 222
 Db 12 DSVKTEVAAPYPSMACTDQGNSSSELTETTRNEIPTTDIKAVEPVKTAGETSAT 71
 QY 223 GTPSIPQNAETITGLKSL---LIGTRDNFTISAEV---DSLLAL----- 262
 Db 72 HTGKREKQLQOMKNNKNDVDNTILSHQKNEFKTKIDTNDSDALLEENQFETNRL 131
 QY 263 -LKESQAPIQGPIQPTLPNNSLATPSLPINPGTSHKEHEDEGYGFANDRIADE 321
 Db 132 HIKOHEVEKDKAKQOKVILKQSD--TKVDSLNDKELNHQKSQAG-----ITNEEK 182
 QY 322 SGVMSHGDSNHYFFKKDTE-----EOLK----- 346
 Db 183 DSNLKIETIRKQAQPDKDAEAVREELKLFSTKAGLDQETQEHVKETSSEWT 242
 QY 347 -----AAQKHEVNTSHGLDLSLSHQ-----DYPGNKEMKDLKKIE 387
 Db 243 QKDEHYANSLQNAQKSL---LDKATNQATQVKNQFLENQAKLKEIQPLIK 295
 QY 388 EKIAQIMQYGVKRESI-----VYKERNALIIYPSGHHHADPIDEKPVG 433
 Db 296 ETNVLYKAMSESLEQVEKELKHNSANEDLVAKSEIYREYEGKLNQSKNLPKLQLE 355
 QY 434 IGHSHNVELFKPEGVAKKEGKRVYTGELTNVNLKSTNNQNTFLANGKRVSF 493
 Db 356 -EAAHSL-----KQVVEOPRKKFKTSEQVTPKRLKRLDANE-----NQOKIELT 402
 QY 494 FPELEKLGINMLVKLITPDGKLVKRVGKVGEGVGNIANFELDOPYLPQGFKTIA 553
 Db 403 VSPE-----NITVE-----GEDVKFTV 421
 QY 554 SKDYPEVSDGTFTVPTSLAYMASQTIFFPHAGDYLVRNPOFVAPVKGTDALVRVEDE 613
 Db 422 AKSDSKTTLD-----FSDLTKYNPSVS-----DRISTN 450
 QY 614 FNGNAYLENNYKVGKILPIPLKNGTTRT-----AGNKIPVTFMANAYLDNOSTVIVE 667
 Db 451 YKNTNT---DNHKAIEITTKNLNSESQVTLKAKDSDSNVVEKTF-----T 493
 QY 668 VPILKENOTDKPSILPQPKRKAQENSKLDEKVEEPTSEKVEKELSETGN----- 720
 Db 494 ITVQKEER-----QVPKTPEDKSDTEKVPQEPKPSNDKQLOBLKSAQOEKLE 545
 QY 721 STSNSTLEVPVY-----DPVQEKVAF-----AESYGHKL 751
 Db 546 ERAIKELMEQPEIPSNPEYGIORSIWESQKEPIQEAITSFKKIIIGDSSSKYTYEHPKY 605
 QY 752 ENVLPMQGTIELYPSGSEVIRKKNMADFTGEAPOGNGENKPSNGKVS-----GTVENQ 806

Db 606 KSHFANYQLHAQM---EMLTRKVVOYMKYIPDPAEIKKIFESDMKRTKEDNNGSLEND 660
 QY 807 PTE-----NK-----PADSLPEADNEKVPKPNSTDNGLMNPENGVGSDPMLD 849
 Db 661 ALKGYFEKYFLTPFKIKQIVDDFDKVEQOQAPPI-PENSE-----MD 703
 QY 850 PALEA-----PAVDPVQEKLEKFTASYGLGL-----DSVIFNMD 884
 Db 704 QAKERAKIAVSKYMSKVLGDGVHQLQKKNHSKIIVDLFELEBAIKQOTIFDID 755

RESULT 14
 US-10-369-493-22154
 Sequence 22154, Application US/10369493
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US 10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 22154
 LENGTH: 1478
 TYPE: PRT
 ORGANISM: Saccharomyces cerevisiae
 US-10-369-493-22154

Query Match 4.28; Score 197; DB 6; Length 1478;
 Best Local Similarity 17.79; Pred. No. 2.2e-05;
 Matches 183; Conservative 139; Mismatches 307; Indels 404; Gaps 47;

QY 34 KDSLSEAEAAAAQAYAKEKGLTPPSTD-----HODSGNTEAKGAEIYNRV-----KAA 82
 Db 251 QEDISNSTSESALSPTKS-GPSKTDKELHSTSTHQTKSASSLYRRSFISLRGSS 309
 QY 83 KKVLDPMYNYQYVEVKNGLI-----IPSYD-----HYHNIRFE----- 119
 Db 310 SNASAKSPSNIKLSIPARPHSIIESNTLTKSAPSPSPSIFRHHKSSSESLL 369
 QY 120 --WEDEGLVEA-----PRGYSLEDLATVYIYVEPRNASDHVRKNKADQSKPDEDEH 171
 Db 370 NSLFGSGIGEAEPTKPNQGHSL-----SSENLAGK-----SKHYETNVS 410
 QY 172 DEVSEPTHPESDEKEN--HAGLNPSADNLYKPSDTEETEEAEEDTDBAEIPGTPSIQ 229
 Db 411 SPLKQSSLTSDDKGLMKNFRKKSQIGVPSNTVAVYTSQE-----TPSLKS 458
 QY 230 NA-----METL-----TGLKSSLLGTDKNNNTISAEVDSLALLKESQ 267
 Db 459 NSSTATLTOTADVNTIPSSSPSPPPPKTANSLEIVISTEDTPKIS-----STASFKETY 514
 QY 268 PAPIQGPQIGQPTLPNNSLATPSLPINPGTSHKEH-----EEDGYGFANDR-- 315
 Db 515 PDCINPK-----IVP-----VFNQKYSVRNELLQDKFPLKKTGLNDSENKVI 560
 QY 316 IIADESGFY-----MSH-----GDSNHYF 335
 Db 561 LVTKDNVSFVPLNLKSVAKLSFESALTKGINHNKNTVHTDFFCDIGAAIPDPTLEF 620
 QY 336 FKDL-----TEQIKAAOKHLEEVKTSUNG--LDSLSSEHQDYPGNKAKMDLD- 383
 Db 621 LKSLFLNTSGKIYKQDMKLOQPKPKAPLTSNNVPLKSVKSSMRSSTSLIATD 680
 QY 384 -----KKIEKIAGIMKQY-----GVRESIYVNEK 410

Db 681 VSIPTSSDITSDHASGSGRRYPQTSYYDRVSNINPTNTEELNWNKE---VLSHEE 737
QY 411 NA--IIPSGDHHADPIDEHK-----PVGIGHSHSNVFLPKPEGVA---KKEGNVY 459
Db 738 NAPKAVFKTSPKLENLDPKSKLNPITPITENESKSFVLRDCEGTIDFNHRRSPY 797
QY 460 TGEEL-----TNVNLKNSFNQNFNFIANGQRKVSFS-----FPPLEKL 502
Db 798 TPPELAPKREAPKPPANTSPORTLSTSK-QNKPRLVRASTRKSRKSKPLPPOL---L 853
QY 503 GINMLVKLITPDGKVLKSVGKVGEGVGNIANFELDPQYLPFGQTFKTIASKDVP--- 558
Db 854 SSPIEASSSPD-----SLTSSITPAST--HVLPQPKGAND 889
QY 559 -----EVSVDGFTVPTSLAYKN-----ASOTIFPPHAGDTYLRVNPQFAVPKG 603
Db 890 VMRLATQDOSTSTSP-SLKMAQKYNRSNSTVSTNSIFYS----- 929
QY 604 TDALVRVDFEFGHAYLENNYKVGKILPDKLNQTTTAGNKIPVTFMANAYLDNQST 663
Db 930 -----PSPLLRGNSKRV---VSSTSAADIFEENDIT 958
QY 664 YIVEVPILEKENTDKPSILPOFKRNKAQENSKLDKVEEPTSEKVEKEKLSGTGNS 723
Db 959 FADAPPNFDSDSDSDSS-----SSDDIWSKKKTAPETNNENKDEKSDNSSTHS 1009
QY 724 NSTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSCGEVKKKNADFTGSA 783
Db 1010 DEIFYDSQTDKMERKT-----FRPSPEVYQNLKFP---F 1043
QY 784 PQNGENKPSNGKVTGIVENOPTENKPADSL-----PEANEPKPKENST 831
Db 1044 PRAN-LDKPITEGAS-----PTSPKSLDLSLSPKNVASSRTEPSTPS-RVPVDSST 1094
QY 832 D-----NGMLNP 838
Db 1095 EFTQDGLNGNKP 1107

RESULT 15
US-09-797-385-6
Sequence 6, Application US/09797385
GENERAL INFORMATION:
APPLICANT: Tal, Joseph Y.
Blake, Milan S.
TITLE OF INVENTION: Non-Iga Fc Binding Forms of the Group B
Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Keseler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/797,385
FILING DATE: 01-Mar-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/923,992
FILING DATE: 05-SEP-1997
APPLICATION NUMBER: US 60/024,707
FILING DATE: (v11)
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-797-385-6
Query Match 4.1%, Score 195.5, DB 5, Length 1128;
Best Local Similarity 17.3%, Pred. No. 1.8e-05;
Matches 154; Conservative 130; Mismatches 296; Indels 309; Gaps 34;
QY 166 DEDKEHDEVSPTHP-----ESDEKENHAGLNPADNLKPSDITDETEEAEDTDAEIP 222
Db 7 DSVKTEVAAPKPPYMAQTDOGNSSSELETTTMEIPTTDIKKAVEPEKTAGETSAT 66
QY 223 GTPSIRQNMETLGLKSSL--LLGTKNNTISAEV-----DSSLAL----- 262
Db 67 DTGKREKQLOQWKNLNDVNTILSHQKNEFKTKIDETNDSALLELENOFNETNRL 126
QY 263 -LKESOPAPIQGPQIGQPTLPN-----NSLATPSPSLPINPQTSHE----- 302
Db 127 HIKQHEEVEKDKKAKQKQKLAQSDTKVDLSNIDKELNKHOKSPVEKMAEPKGITNEDKDSM 186
QY 303 -----KHEEDGVGFPAKRIIADESGFVMSGDSN-----HYFFKOLTEQ----- 344
Db 187 LKKIEDIRKQAQADKEDAEVYREELGKLFSSKAGLDQEIHEHVKKTSSEENTQKV 246
QY 345 -----IKAAQHLEEVKTSNGLDLSLSEHQ-----DYPGNAKEMKDLKKEIKI 390
Db 247 DEHYANSQNLAKSLEE-----LDKATNEQATQVKNQFLENAOKLKEQOPLIKETN 299
QY 391 AGIMKQYGVKRESI-----VYKKEKNAIIPSGDHHADPIDEHRPVGIGH 436
Db 300 VKLYKAMSESLQVEKELKHSEANLEDLVAKSKEIVREYEGKLNQKNLPKOLE-EE 358
QY 437 SHSNYELFKPEGVAKKEGKVTYTGELTNVYVLLKNSTFNNOFTLANGOKRVFSFPP 496
Db 359 AHSKL-----KOVVEDFRKFKTKTSEQVTPKKRVARDLAANE-----NNOOKIELTSP 406
QY 497 ELEKGLGINMLVKLITPDGKVLKSVGKVGEGVGNIANFELDPQYLPFGQTFKTIASKD 556
Db 407 E-----NITVYE-----GEDVKFTVTAKS 425
QY 557 YPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGIDALVRFVDFEFG 616
Db 426 DSKTTLD-----FSDLLTKYNPSVS-----DRISTNTYKT 454
QY 617 NAYLENNYKVGKILPKLNQTTT-----AGNKIPVTFMANAYLDNQSTYIVEVPI 670
Db 455 NT---DNHKAIEITIKNLKLNESQTVTLAKDSDGNNVEKTP-----TIV 497
QY 671 LEKENTDKPSILPOFKRNKAQENSKLDKVEEPTSEKVEKEKLSGTG-----STS 723
Db 498 QKKEE-----QVPKTPQKDSKTEKVPQEPKSDKNQLOELIKQAQOAELEKLEKA 549
QY 724 NSTLEEVPTV-----DPVQEKVAKF-----AESYGMKLENV 754
Db 550 IKLEMEQPEIPSNPEYGIKSIWESQKEPIQEAITFKKLIIGDSSSKYTYEYFNKSKD 609
QY 755 LFNMDGTIELYLPSCGEVIRKKNMADFTCEAPQNGENKPSNGKVS-----GVENQOPT 809
Db 610 FMYQLHAQN-----EMLTRKVVQYNNKYPDIAEIKKIFESDMKRTKEDNYGSLNDALK 664
QY 810 -----NK-----PADSLPEAPNEKVPKPNSTNDGMLNPEGNVGSFMDLPAL 852
Db 665 GYFEKVELTFFNKIKQIVDDDKKVEQDQAPI-PENSE-----MDQAK 707
QY 853 EEA-----PAVDPVQEKLEKFTASYGLGL-----DSVIFNMD 884

9:21:11 2003

' : | : | | : | : | | : | || :: |
EKAKIAVSKYMSKVLGDGVHQHQLQKKNHHSKIVDLFKELEAIKQOTIFDID 756

5. .sch completed: March 27, 2003, 12:04:41
Job time : 56 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:53:08 ; Search time 43 Seconds
(without alignments)
2807.560 Million cell updates/sec

Title: US-09-884-465a-332
Perfect score: 4728
Sequence: 1 MQTYTDEQVAKLAKYK.....IELRLPSGEVKKNSLDFIA 906

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID82/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID82/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	4728	100.0	907	23	AAU84065 S. pneumoniae deri
2	4723	99.9	907	23	AAU84061 S. pneumoniae deri
3	4705	99.5	907	23	AAU84063 S. pneumoniae deri
4	4675	98.9	901	23	AAU84067 S. pneumoniae deri
5	4672	98.8	901	23	AAU84068 S. pneumoniae deri
6	4663	98.6	901	23	AAU84072 S. pneumoniae deri
7	4663	98.6	901	23	AAU84074 S. pneumoniae deri
8	4656.5	98.5	902	23	AAU84076 S. pneumoniae deri
9	4652.5	98.4	902	23	AAU84078 S. pneumoniae deri
10	4624	97.8	895	23	AAU84069 S. pneumoniae deri

11	4619	97.7	895	23	AAU84070 S. pneumoniae deri
12	4615	97.6	895	23	AAU84073 S. pneumoniae deri
13	4615	97.6	895	23	AAU84075 S. pneumoniae deri
14	4608.5	97.5	896	23	AAU84077 S. pneumoniae deri
15	4608.5	97.5	896	23	AAU84080 S. pneumoniae deri
16	4604.5	97.4	896	23	AAU84079 S. pneumoniae deri
17	4571	96.7	889	23	AAU84071 S. pneumoniae deri
18	4560.5	96.5	890	23	AAU84081 S. pneumoniae deri
19	4518	95.6	1152	23	AAU84054 S. pneumoniae deri
20	4518	95.6	1378	23	AAU84053 S. pneumoniae deri
21	4229.5	89.5	913	23	AAU84059 S. pneumoniae deri
22	4229.5	89.5	1139	23	AAU84055 S. pneumoniae deri
23	4147	87.7	1365	23	AAU84057 S. pneumoniae deri
24	4014.5	84.9	1238	23	AAU84056 S. pneumoniae deri
25	3848.5	81.4	1136	23	AAU84058 S. pneumoniae deri
26	3596	76.1	999	23	AAU84051 S. pneumoniae deri
27	3375	71.4	1019	21	AAU84051 Streptococcus pneu
28	3375	71.4	1019	21	AAU84051 Streptococcus pneu
29	3375	71.4	1019	21	AAU84051 Streptococcus pneu
30	3375	71.4	1039	21	AAU84021 Streptococcus pneu
31	3375	71.4	1039	21	AAU84021 Streptococcus pneu
32	3359	71.0	1019	21	AAU84021 Streptococcus pneu
33	3359	71.0	1019	21	AAU84021 Streptococcus pneu
34	3359	71.0	1019	21	AAU84021 Streptococcus pneu
35	3357	71.0	1019	21	AAU84021 Streptococcus pneu
36	3356	71.0	1019	21	AAU84021 Streptococcus pneu
37	3300.5	69.8	807	23	AAU83997 Streptococcus pneu
38	3298	69.8	632	23	AAU84009 Streptococcus pneu
39	3298	69.8	901	23	AAU84056 Streptococcus pneu
40	3298	69.8	907	23	AAU84060 Streptococcus pneu
41	3298	69.8	907	23	AAU84084 Streptococcus pneu
42	3293	69.6	632	23	AAU84006 Streptococcus pneu
43	3293	69.6	632	23	AAU84013 Streptococcus pneu
44	3288	69.5	632	23	AAU83998 Streptococcus pneu
45	3288	69.5	644	23	AAU83996 Streptococcus pneu

ALIGNMENTS

RESULT 1
AAU84065
ID AAU84065 standard; Peptide; 907 AA.
XX
AC AAU84065;
DT 08-MAY-2002 (first entry)
DE S. pneumoniae derived chimeric peptide, VP94.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumonia; streptococcal bacterial infection; mutant; mutain;
XX BVH-11-2.
XX Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
epitope-bearing polypeptides, useful as vaccine components for treating

enting streptococcal infections such as otitis media,
meningitis, and bacteraemia

Claim 1; Page -: 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 907 AA:

Query Match	100.0%	Score 4728;	DB 23;	Length 907;
Best Local Similarity	100.0%	Pred. No. 6.1e-275;		
Matches 906;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	MQITYTDEIQVAKLAGKTYTDEGYIFDTSWIKKDSLSAEARAAQAYAKERGLTPPSTD	60		
DB 2	MQITYTDEIQVAKLAGKTYTDEGYIFDTSWIKKDSLSAEARAAQAYAKERGLTPPSTD	61		
QY 61	QDSGNTAEKGAIAIYRVKAAKVPDLPYNIQYTVFVNGSLIIPSYDHYHNKFEW	120		
DB 62	QDSGNTAEKGAIAIYRVKAAKVPDLPYNIQYTVFVNGSLIIPSYDHYHNKFEW	121		
QY 121	FDEGLYAPKGYSLIEDLATYKYIYVPRNASDHRVKNKADQSDKPEDKEHDEVSSEPTHP	180		
DB 122	FDEGLYAPKGYSLIEDLATYKYIYVPRNASDHRVKNKADQSDKPEDKEHDEVSSEPTHP	181		
QY 181	ESDEKENHAGLNPSAOLYKPSDTTETEEDTDEAEIPGTPSIQONAMETLTGLKS	240		
DB 182	ESDEKENHAGLNPSAOLYKPSDTTETEEDTDEAEIPGTPSIQONAMETLTGLKS	241		
QY 241	SLILGTRDNTTISAEVDSLALAKESQAPIQGPTLPNNSLATPSPSLPINFPTS	300		
DB 242	SLILGTRDNTTISAEVDSLALAKESQAPIQGPTLPNNSLATPSPSLPINFPTS	301		
QY 301	HEKHEEDGYGFDANRIIAEDSGFVMSHSDSNHYFFKKDLTEBQIKAAQKHLEEVKTSN	360		
DB 302	HEKHEEDGYGFDANRIIAEDSGFVMSHSDSNHYFFKKDLTEBQIKAAQKHLEEVKTSN	361		
QY 361	GLDSLSEHQDYPONAKEMDLDKIEETAGIMQYGVKRESIVYNKKNATIIYPSGDH	420		
DB 362	GLDSLSEHQDYPONAKEMDLDKIEETAGIMQYGVKRESIVYNKKNATIIYPSGDH	421		
QY 421	HHADPIDEKHPVGIHSHSNHYELFKPEGVAKEGKNVYTGELTNNVLLKNSTFNQ	480		
DB 422	HHADPIDEKHPVGIHSHSNHYELFKPEGVAKEGKNVYTGELTNNVLLKNSTFNQ	481		
QY 481	FTLANGKRVSFSPPELEKLGIMLVKLITPDGKVLKRVKSGVGEVGNANTFELQ	540		
DB 482	FTLANGKRVSFSPPELEKLGIMLVKLITPDGKVLKRVKSGVGEVGNANTFELQ	541		
QY 541	PYLPQOTFKYTIASKDYPEVSDGTFTVPTSLAYKMASQITFTFPFHAGDTYLRVNPQFV	600		
DB 542	PYLPQOTFKYTIASKDYPEVSDGTFTVPTSLAYKMASQITFTFPFHAGDTYLRVNPQFV	601		

QY	601	PKGTDAIVRVDEFHGNAYLENNYKVGTEIKLPKLNQCTTTRTAGNKIPVTMANAYLDN	660
DB	602	PKGTDAIVRVDEFHGNAYLENNYKVGTEIKLPKLNQCTTTRTAGNKIPVTMANAYLDN	661
QY	661	QSTYIIVEVPILEKENOTDKPSILPQFKRKAQNSKLDKVEEPTSEKVEKEKLSGTGN	720
DB	662	QSTYIIVEVPILEKENOTDKPSILPQFKRKAQNSKLDKVEEPTSEKVEKEKLSGTGN	721
QY	721	STNSSTLEEVPTVDPQVKAKFAESYGKLENVLFNMDGTIELYLPSPGVIKKNAWDT	780
DB	722	STNSSTLEEVPTVDPQVKAKFAESYGKLENVLFNMDGTIELYLPSPGVIKKNAWDT	781
QY	781	GEAPOGNGENKPSGKSTGTVENOPTENKPADSLPEAPNEKPVKPESTONGMLNPEG	840
DB	782	GEAPOGNGENKPSGKSTGTVENOPTENKPADSLPEAPNEKPVKPESTONGMLNPEG	841
QY	841	NVGSDDPMLDPALEAPAVDPQVKLEKFTASYGLDSDVFNMDGTIELYLPSPGVIKKN	900
DB	842	NVGSDDPMLDPALEAPAVDPQVKLEKFTASYGLDSDVFNMDGTIELYLPSPGVIKKN	901
QY	901	LSDFIA 906	
DB	902	LSDFIA 907	

RESULT 2

AAU84061

ID: AAU84061 standard; Peptide; 907 AA.

XX AC AAU84061;

DT 08-MAY-2002 (first entry)

DE S. pneumoniae derived chimeric peptide, VP90.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumoniae; streptococcal bacterial infection; mutant; mutain;
XX BVH-11-2.

OS Streptococcus pneumoniae.
OS Synthetic.

XX WO200198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA00908.

XX 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/16.

PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

PS Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (iii) encoding (i) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (iii) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Query Match 99.9%; Score 4723; DB 23; Length 907;
Best Local Similarity 99.9%; Pred. No. 1.2e-274; Indels 0; Gaps 0;
Matches 905; Conservative 0; Mismatches 1;

QY 1 MOITTTDDEIOVAKLAGYTTEDGYIFTDSWIKKDSLSAEARAAQAYAKEGLTPPSTD 60
DB 2 MOITTTDDEIOVAKLAGYTTEDGYIFTDSWIKKDSLSAEARAAQAYAKEGLTPPSTD 61
QY 61 HQDSGNTAKGAEATYNNRKAAYKVPLOMPYNTQYTVVEVNGSLIIPSYDHYHNKFEW 120
DB 62 HQDSGNTAKGAEATYNNRKAAYKVPLOMPYNTQYTVVEVNGSLIIPSYDHYHNKFEW 121
QY 121 FDEGLYAPKGSLELLATVYKVEPRNADSVHRKNADQSKPDEKDEHVESEPH 180
DB 122 FDEGLYAPKGSLELLATVYKVEPRNADSVHRKNADQSKPDEKDEHVESEPH 181
QY 181 ESDEKENHAGLNPADNLYKPKSTDETEEEAEDTDEAIPGTPSIRQNAETITGLKS 240
DB 182 ESDEKENHAGLNPADNLYKPKSTDETEEEAEDTDEAIPGTPSIRQNAETITGLKS 241
QY 241 SLLGTYKDNNTISAEDVSLALLKESQAPIQGPTLPNNSLATPSPSLNPCTS 300
DB 242 SLLGTYKDNNTISAEDVSLALLKESQAPIQGPTLPNNSLATPSPSLNPCTS 301
QY 301 HEKHEEDGFGDANRIIADESGFVNSHSDSNHYFPKDLTEQIKAAQKHEEVKTSN 360
DB 302 HEKHEEDGFGDANRIIADESGFVNSHSDSNHYFPKDLTEQIKAAQKHEEVKTSN 361
QY 361 GLDLSLSEHDQYPCNAKEMKDKLTKTEKAGIMKQYGVKRSIVVVKENAIIPSGDH 420
DB 362 GLDLSLSEHDQYPCNAKEMKDKLTKTEKAGIMKQYGVKRSIVVVKENAIIPSGDH 421
QY 421 HHADPTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNTTGEELTNVLLKNSTFNQ 480
DB 422 HHADPTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNTTGEELTNVLLKNSTFNQ 481
QY 481 FTLANGKRVSFPPPELEKLGINMLVLITPDGKVLKSVKGVGEGVGNIANFELQ 540
DB 482 FTLANGKRVSFPPPELEKLGINMLVLITPDGKVLKSVKGVGEGVGNIANFELQ 541
QY 541 PYPGQFTKTIASKDYFVSYGTTFTVPSLAYKASQTIFPFFHAGTYLRVNPQFV 600
DB 542 PYPGQFTKTIASKDYFVSYGTTFTVPSLAYKASQTIFPFFHAGTYLRVNPQFV 601
QY 601 PKGTDLRVFDEPHGNAYLNNYKVEIKLPKLNQGTTRTAGNKIPVTFMANAYLON 660
DB 602 PKGTDLRVFDEPHGNAYLNNYKVEIKLPKLNQGTTRTAGNKIPVTFMANAYLON 661
QY 661 QSTYIVVEPILKENQTDKPSILPQFRKNAQENSKLDEKVEPKTSERKEKELSETGN 720
DB 662 QSTYIVVEPILKENQTDKPSILPQFRKNAQENSKLDEKVEPKTSERKEKELSETGN 721
QY 721 STNSLTLEVTVPQVOKVAKFAESGKLENLVFNMDGTIELYLSGGEVVKMAADFT 780
DB 722 STNSLTLEVTVPQVOKVAKFAESGKLENLVFNMDGTIELYLSGGEVVKMAADFT 781
QY 781 GEAPGNGENKPSNGKSTGTVENQPTENKPADSLPEAPNEKPKPENSTNDGMLNPEG 840

DB 782 GEAPGNGENKPSNGKSTGTVENQPTENKPADSLPEAPNEKPKPENSTNDGMLNPEG 841
QY 841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDCGTIELRLPSPGVIKKN 900
DB 842 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDCGTIELRLPSPGVIKKN 901
QY 901 LSDFIA 906
DB 902 LSDFIA 907
RESULT 3
AAU84063
ID AAU84063 standard; Peptide; 907 AA.
XX
AC AAU84063;
XX
DT 08-MAY-2002 (first entry)
DE S. pneumoniae derived chimeric peptide, VP92.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WC200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA09098.
XX
PR 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
XX
XX Claim 1; Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (i) with 70-90A
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (ii)
XX comprising (i) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (ii) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
XX Staphylococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (iii) encoding (i) is useful in a diagnostic
XX techniques. The Streptococcus polypeptides are useful for designing DNA
XX test for S. pneumoniae infection. (iii) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumoniae nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a chimeric
XX gene created from fragments and variant fragments of Streptococcus
XX pneumoniae genes, described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX

ID	AAU84067	standard; Peptide; 901 AA.
XX	AAU84067;	
AC	08-MAY-2002	(first entry)
XX	S. pneumoniae	derived chimeric peptide, VP109.
DT	BVH-3; BVH-11;	vaccine; meningitis; otitis media; bacteraemia;
XX	pneumonia;	streptococcal bacterial infection; mutant; mutein;
DE	BVH-11-2.	
XX	Streptococcus	pneumoniae.
OS	Synthetic.	
XX	WO200198334-A2.	
PN	27-DEC-2001.	
XX	19-JUN-2001;	2001WO-CA00908.
XX	20-JUN-2000;	2000US-212683P.
PF	(SHIR-) SHIRE	BIOCHEM INC.
XX	Hamel J, Ouellet C,	Charland N, Martin D, Brodeur B;
PI	WPI; 2002-12272/16.	
DR	New Streptococcus	pneumoniae BVH-3 and BVH-11 variant and
XX	epitope-bearing	polypeptides, useful as vaccine components for treating
PT	or preventing	streptococcal infections such as otitis media,
PT	meningitis,	and bacteraemia
XX	Claim 1; Page -;	113pp; English.
PS	The invention	describes an isolated polypeptide (I) with 70-90%
XX	identity to	Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
CC	BVH-3 or BVH-11,	or chimeric sequences derived from them. A vaccine (II)
CC	comprising (I) is	useful for therapeutic or prophylactic treatment of
CC	meningitis, otitis	media, bacteraemia or pneumonia infection in an
CC	individual	susceptible to these disorders. (II) is also useful for
CC	therapeutic or	prophylactic treatment of any streptococcal bacterial
CC	infection (e.g.,	caused by Streptococcus pneumoniae, group A
CC	Streptococcus	such as Streptococcus pyogenes, group B Streptococcus or
CC	Streptococcus	agalactiae, S. dysgalactiae, S. uberis, S. noxia or
CC	Staphylococcus	aureus) in an individual susceptible to the infection.
CC	A polynucleotide	(iii) encoding (I) is useful in DNA immunisation
CC	techniques. The	Streptococcus polypeptides are useful in a diagnostic
CC	test for S.	pneumoniae infection. (iii) is useful for designing DNA
CC	probes for use	in detecting the presence of Streptococcus in a biological
CC	sample suspected	of containing the bacteria. The DNA probes may also be
CC	used for detecting	circulating S. pneumoniae nucleic acid in a sample for
CC	diagnosing	streptococcal infections. This sequence represents a chimeric
CC	gene created	from fragments and variant fragments of Streptococcus
CC	pneumoniae	genes, described in the method of the invention.
CC	Note: This	sequence does not appear in the specification but has
CC	been created	according to information given in the invention.
XX	Sequence	901 AA;
QY	Query Match	98.9%; Score 4675; DB:23; Length 901;
DB	Best Local Similarity	99.3%; Pred. No. 9,1e-273;
DB	Matches 900;	Conservative 0; Mismatches 0; Indels 6; Gaps 1;
QY	1	MQITTTDDDEIQVAKLAGKTYTTEGDIYFTTSWIKKDSLEAERAAQAAYAKKGLTPPSTD 60
DB	2	MQITTTDDDEIQVAKLAGKTYTTEGDIYFTTSWIKKDSLEAERAAQAAYAKKGLTPPSTD 61
QY	61	HQDSGTEAKGAEATYNNRVKAAKVPDLPMPYNNQYTVVEKNGSLIIPSYDHYHNKFEW 120
DB	62	HQDSGTEAKGAEATYNNRVKAAKVPDLPMPYNNQYTVVEKNGSLIIPSYDHYHNKFEW 121
QY	121	FDGLYEAPKGSLEDLATVYIYVEPRNASDVRKADQSKPDEKDEHDEVESEPH 180
DB	122	FDGLYEAPKGSLEDLATVYIYVEPRNASDVRKADQSKPDEKDEHDEVESEPH 181
QY	181	ESDEKENHAGLNPADNLYKPTDTEETEEAEADTDEAETPGTPIRQNAWETLTGLKS 240
DB	182	ESDEKENHAGLNPADNLYKPTDTEETEEAEADTDEAETPGTPIRQNAWETLTGLKS 241
QY	241	SLLLGTNDNTISAEDVSLALLKESQAPAIQGIQPTLPNNSLATPSPSLINPCTS 300
DB	242	SLLLGTNDNTISAEDVSLALLKESQAPAIQGIQPTLPNNSLATPSPSLINPCTS 301
QY	301	HEKHEEDGYGDFDANRIIAEDSGFVMSHGDSNHYFFKDLTTEQIKAAQKHLEEVKTSN 360
DB	302	HEKHEEDGYGDFDANRIIAEDSGFVMSHGDSNHYFFKDLTTEQIKAAQKHLEEVKTSN 361
QY	361	GLDLSHEDQYDGNKAKMDLKKTEEKIAGIMKQYGVKRESIVVYKKNAIIPSGDH 420
DB	362	GLDLSHEDQYDGNKAKMDLKKTEEKIAGIMKQYGVKRESIVVYKKNAIIPSGDH 421
QY	421	HHADPIDEHRPVGIGHSHNYELFKPEGVAKKEGKNTYTGELTNVNLKNSTFNQ 480
DB	422	HHADPIDEHRPVGIGHSHNYELFKPEGVAKKEGKNTYTGELTNVNLKNSTFNQ 481
QY	481	FTLANGKRVSPFPPELEKGLINMLVLTDPGKVLKVGKVGEGVGNIANFELQ 540
DB	482	FTLANGKRVSPFPPELEKGLINMLVLTDPGKVLKVGKVGEGVGNIANFELQ 541
QY	541	PYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLNVPQFV 600
DB	542	PYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLNVPQFV 601
QY	601	PKGTDLVRVDFDEPHGNAYLNNYKVEIKLPIPKLNOGTTRAGNKIPVTFMANAYLON 660
DB	602	PKGTDLVRVDFDEPHGNAYLNNYKVEIKLPIPKLNOGTTRAGNKIPVTFMANAYLON 661
QY	661	QSTVIPEVPILEKENQDKPSILLPOPKRKAQNSKLDKVEEPKTSKVEKEKLSGTN 720
DB	662	QSTVIPEVPILEKENQDKPSILLPOPKRKAQNSKLDKVEEPKTSKVEKEKLSGTN 721
QY	721	STNSNTEEVTPDVPQEKVAKFAESTGMKLENVFNQGTIELIYLPSPGVIKKNADFT 780
DB	722	STNSNTEEVTPDVPQEKVAKFAESTGMKLENVFNQGTIELIYLPSPGVIKKNADFT 781
QY	781	GEAPGNGENKPSNGKYSTGTSTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEG 840
DB	782	GEAPGNGENKPSNGKYSTGTSTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEG 841
QY	841	NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDDGTIELRLPSPGVIKKN 900
DB	842	NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDDGTIELRLPSPGVIKKN 901
QY	901	LSDFTA 906
DB	902	LSDFTA 907
RESULT 4		
AAU84067		

Fri Mar 28 09:21:09 2003

QY 121 FDEGLYEPKGYSLDILATVYKVVPRNASHVRKNKADQDSKDPDEKHEHDEVSEPTHP 180
Db 116 FDEGLYEPKGYSLDILATVYKVVPRNASHVRKNKADQDSKDPDEKHEHDEVSEPTHP 175
QY 181 ESDEKENHAGLNPSADNLYKPSDTTETEEAEEDTDEAIPOTPSIRQNAEMETLGLKS 240
Db 176 ESDEKENHAGLNPSADNLYKPSDTTETEEAEEDTDEAIPOTPSIRQNAEMETLGLKS 235
QY 241 SLLGTDKNTTISAEVDSLLALKESQAPIQGPTLPNNSLATPSPSLPINFPTS 300
Db 236 SLLGTDKNTTISAEVDSLLALKESQAPIQGPTLPNNSLATPSPSLPINFPTS 295
QY 301 HEKHEEDYGFDAENRIIAEDSGFVMSHGSDNHYFFKKDLTEEQIKAAQKHLEEVKTSN 360
Db 296 HEKHEEDYGFDAENRIIAEDSGFVMSHGSDNHYFFKKDLTEEQIKAAQKHLEEVKTSN 355
QY 361 GLDSSSHQDYPGNAEKMDLKIIEKTAGIMQYGVKRESIVVKNENAIIPSGDH 420
Db 356 GLDSSSHQDYPGNAEKMDLKIIEKTAGIMQYGVKRESIVVKNENAIIPSGDH 415
QY 421 HADPIDHKKVPGIGHSHSNYELEKPEEGVAKKEGNKYTTGEELTNVNLKSTFNQN 480
Db 416 HADPIDHKKVPGIGHSHSNYELEKPEEGVAKKEGNKYTTGEELTNVNLKSTFNQN 475
QY 481 FTLANGQRVSFPPELEKLGIMLVKLITPDGKVLKESGVFGVGVGNIANFELDQ 540
Db 476 FTLANGQRVSFPPELEKLGIMLVKLITPDGKVLKESGVFGVGVGNIANFELDQ 535
QY 541 PYLPQGTFTYITASKDYPEVSDGFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 600
Db 536 PYLPQGTFTYITASKDYPEVSDGFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 595
QY 601 PRGTDALRVDFEFGHAGNLYENNYKVGELKIPKLNGQTTNAGNKIPVTFMANAYLDN 660
Db 596 PRGTDALRVDFEFGHAGNLYENNYKVGELKIPKLNGQTTNAGNKIPVTFMANAYLDN 655
QY 661 QSYIYVEVILEKENQDTPSILPOKRNKAQENSKLDEKVEEPTSEKVEKLSGTN 720
Db 656 QSYIYVEVILEKENQDTPSILPOKRNKAQENSKLDEKVEEPTSEKVEKLSGTN 715
QY 721 STNSNLEEVPTDVPQEVAKFAESYGMKLENVLFNMDGTIELYLPSEGVKKMAOFT 780
Db 716 STNSNLEEVPTDVPQEVAKFAESYGMKLENVLFNMDGTIELYLPSEGVKKMAOFT 775
QY 781 GEAPQNGENKPSNGKSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNGMLNPEG 840
Db 776 GEAPQNGENKPSNGKSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNGMLNPEG 835
QY 841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRPSGEVIKKN 900
Db 836 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRPSGEVIKKN 895
QY 901 LSDPIA 906
Db 896 LSDPIA 901
RESULT 5
AAU84068
ID AAU84068 standard; Peptide; 901 AA.
XX
XX AAU84068;
AC
XX 08-MAY-2002 (first entry)
DT
XX
XX S. pneumoniae derived chimeric peptide, VP110.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX Streptococcus pneumoniae.
OS Synthetic.

XX W0200198334-A2.
PN
XX 27-DEC-2001.
PD
XX 19-JUN-2001; 2001WO-CA00908.
PF
XX 20-JUN-2000; 2000US-212683P.
PR
XX (SHIR-) SHIRE BIOCHEM INC.
PA
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI; 2002-122272/16.
XX
DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
PT
XX
PS Claim 1; Page -; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX Sequence 901 AA;
QY
Query Match 98.8%; Score 4672; DB 23; Length 901;
Best Local Similarity 99.2%; Pred. No. 1.4e-271;
Matches 899; Conservative 0; Mismatches 1; Indels 6; Gaps 1;
QY 1 MQITVTDDEIOVAKLAGYTTEDGYIFDTSWIKKDSLSAEARAAQAYAKEKGLTPPSTD 60
Db 2 MQITVTDDEIOVAKLAGYTTEDGYIFDTSWIKKDSLSAEARAAQAYAKEKGLTPPSTD 61
QY 61 HDSCNTEAKGAEATYNEVKAAKVPLDRMPYNLQYTVVEKNGSLIIPSYDHYHNKFEW 120
Db 62 HDSCNTEAKGAEATYNEVKAAKVPLDRMPYNLQYTVVEKNGSLIIPSYDHYHNKFEW 121
QY 121 FDEGLYEPKGYSLDILATVYKVVPRNASHVRKNKADQDSKDPDEKHEHDEVSEPTHP 180
Db 122 FDEGLYEPKGYSLDILATVYKVVPRNASHVRKNKADQDSKDPDEKHEHDEVSEPTHP 181
QY 181 ESDEKENHAGLNPSADNLYKPSDTTETEEAEEDTDEAIPOTPSIRQNAEMETLGLKS 240
Db 182 ESDEKENHAGLNPSADNLYKPSDTTETEEAEEDTDEAIPOTPSIRQNAEMETLGLKS 241
QY 241 SLLGTDKNTTISAEVDSLLALKESQAPIQGPTLPNNSLATPSPSLPINFPTS 300
Db 242 SLLGTDKNTTISAEVDSLLALKESQAPIQGPTLPNNSLATPSPSLPINFPTS 301
QY 301 HEKHEEDYGFDAENRIIAEDSGFVMSHGSDNHYFFKKDLTEEQIKAAQKHLEEVKTSN 360
Db

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia

Claim 1; Page -: 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae, S. pneumoniae aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 901 AA;

Query Match 98.6%; Score 4563; DB 23; Length 901;
Best Local Similarity 99.1%; Pred. NO. 4.8e-271;
Matches 898; Conservative 0; Mismatches 2; Indels 6; Gaps

1 MQIYTTDEIQAVKLAGKTTEDGYIFDTSWIKKDSLSAEARAAQAYAKEGLTPPSTD 60
2 MQIYTTDEIQAVKLAGKTTEDGYIFDTSWIKKDSLSAEARAAQAYAKEGLTPPSTD 61
61 HQDSGNTAEKGAEAIIYNRVKAARKVPLDRMPNYLQYTVVEKNGSLIIPSYDHYINIKFEW 120
62 HQDSGNTAEKGAEAIIYNRVKAARKVPLDRMPNYLQYTVVEKNGSLIIPSYDHYINIKFEW 121
121 FDEGLYEPAPGYSLIEDLLATYKYTYVEPRASDHVRNKAADQSKDEKDEHVESETHP 180
122 FDEGLYEPAPGYSLIEDLLATYKYTYVEPRASDHVRNKAADQSKDEKDEHVESETHP 181
181 ESDKENHAGLNFSADNLKYPSTDTEETEEEAEDTTDEAETPGTSPSRONAMETLTGLKS 240
182 ESDKENHAGLNFSADNLKYPSTDTEETEEEAEDTTDEAETPGTSPSRONAMETLTGLKS 241
241 SLLGLTKDNTTISAEVDSLALALKESQAPIQGPQIGQPTLPNNSLATPSPSLINPGTS 300
242 SLLGLTKDNTTISAEVDSLALALKESQAPIQGPQIGQPTLPNNSLATPSPSLINPGTS 301
301 HEKHEEDGYGFDANRIIATDESQFVMSHSDNHYHFFPKDLTEEQIKAAQKHLKEVKTSHN 360
302 HEKHEEDGYGFDANRIIATDESQFVMS -----YFFKKDLTEQIKAAQKHLKEVKTSHN 355
361 GLDLSLSHEQDYPGNAKEMKDLKKIEEKIAGIMKQYGVKRESIVVYVNEKNAIYPSGDH 420
362 GLDLSLSHEQDYPGNAKEMKDLKKIEEKIAGIMKQYGVKRESIVVYVNEKNAIYPSGDH 421
421 HHAADPIDHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVYVNEKNAIYPSGDH 480
422 HHAADPIDHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVYVNEKNAIYPSGDH 481
481 FTLANGQRVRSFPPELEKLGINMLVLIITPDGKVLKVSCKVFGEGVGHIANFELDQ 540
482 FTLANGQRVRSFPPELEKLGINMLVLIITPDGKVLKVSCKVFGEGVGHIANFELDQ 541
541 FTLANGQRVRSFPPELEKLGINMLVLIITPDGKVLKVSCKVFGEGVGHIANFELDQ 542
542 FTLANGQRVRSFPPELEKLGINMLVLIITPDGKVLKVSCKVFGEGVGHIANFELDQ 543

CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

AA	Sequence	902 AA;
SQ		

Query Match 98.5%; Score 4656.5; DB 23; Length 902;
Best Local Similarity 99.0%; Pred. No. 1.2e-270;
Matches 898: Conservative 0; Mismatches 2; Indels 7;

QY 1 MQITYDDEIQVAKLACKYTTEDGYIFDTSWIKKDSLSEARAAQAYAKEGLTPSTD 60

61 HQDSGTEAKGAEEIYNRVKAARKVPDRMPYNLOITVEVKNGSLIPSYDHYHNKIFEW
QY

Db 62 HQDSGNTKAGAAEAIYNRVKAAKKVPLDRMPYNLTQTVVEKNGSLIIPHYDVHYHNKFEW 122

121	FDEGLYEAPKGSYLEDLLATVKVYVVEPRNASDRHYRNKKAQDSKPPDEKHEHDEVSEPTHP	18
QY		19
		20
		21
		22
		23
		24
		25
		26
		27
		28
		29
		30
		31
		32
		33
		34
		35
		36
		37
		38
		39
		40
		41
		42
		43
		44
		45
		46
		47
		48
		49
		50
		51
		52
		53
		54
		55
		56
		57
		58
		59
		60
		61
		62
		63
		64
		65
		66
		67
		68
		69
		70
		71
		72
		73
		74
		75
		76
		77
		78
		79
		80
		81
		82
		83
		84
		85
		86
		87
		88
		89
		90
		91
		92
		93
		94
		95
		96
		97
		98
		99
		100

Db 122 FDGLYEAPKGYSLDLLATVKYIVBPRNASDVRKRNKABQSKPDEDKHEDEVSEPTHP 10

[illegible]

Db 182 ESDEKENHAGLNPSADNLYKPSDTEETEEAEEDTDEAEIPGTPSPSRQNAMEYLAGKS 21

241 SLLLGTKNNNTISA EVDLSUALLKESPAPIQ-GPQIGQPTLPNNNSLAFSPSPSLP INPGT 301

```

242 SLLIGTKONTISA EVDLSLLALLKRESPAPIQSGRQIGQPTLNNSLAFSPSLP INPGT
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300

```

	QY	300 SHEKHEEDGYGFDANRIIAEDESFGVASHGDSNHFFFKRLPEEQIKAAQKHLEEVATISH	35
--	----	--	----

Db 302 SHEKHEEDGYGFDANRIIADESGFVMS-----YFFKRLDTEEQINAKQKHLLEAVATSH 303

360 NGDLSLSHEQDPGNAKENKOLDKRIEERIKAGIKQVGRRESIWNKEKNALLIPSGD 411

[illegible]

Qy 420 HHADPIDEHKPVGIGHSHSNVELFKPEGVAKKGNKVTGEELTNVVNLLKNSTFNNQ 47

Db 416 HHHADPIDEHKPGVIGHSHSNYELFKPEGVAKKECNKVTGELTNVNVLLKNSTFNNQ 47

480 NFTLANGQKRVSFPPELEKKLGINLVKLITPDGKVLKVSCKVFGEVGNIANFELD 53

Db 476 NFTLANGQKRVSFPPELEKKGINMLVKLITPDGKVLKVSQKVGEGVGNIAFELD 53

Qy 540 QPYLPQGTFFKYTIASKDYPEVSYDGTFFVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFA 59

Db 536 QPYLPQTEKYYTIASKDYPEVSYDGTFTVPTSLAYKNASQTIFFPFHAGDTYLRVNPQFA 59

Qy 600 VPKGTDALVRVDFEHGHNAYLENNYKVGSEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLD 65

Db 596 VPKGTDLVRVDFEFHGNAYLENNYKVGEEKLPIPKLNQGTTRTAGNKKIPVTFMANAYLD 65

QY 660 NQSTYIVPILKENQTKPSILPQFRNKAQENSKLDEKVEBPKTSEKYKEKLSGT 71

Db 656 NQSTYIYVEVPILKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTKSEKVEKELSETG 71

QY 720 NSTSNSTLEEPTVDPPQEKVAKFAESYGMKLENTLFNMDGTIELYLPSGETVKKNNDAF 77

Db 716 NSTSNSTLEEPTVDPVQEKVAKFAESYGMKLENTLFPNMDGTIELYLPSGEVIKKNMADE 77

Qy 780 TGEAPOGNGENKPSBNGKVGSTGTVENQPTENKPADSLPEAPNEAPVKPENSTNDGMLNPE 83

Db 776 TGEAPQNGENKPSGKVGSTGTVENQPTENKPADSLPEAPNEAPVKPENSDNGMLNPE 83

Qy 840 GNVGSDPMLDPALEAPAVDPVQEKLEKTASYGLGLDSVFNMDGTIELRLPSPGEVIKK 89

Db 836 GNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKK 899

Qy 900 NLSDFIA 906
|||||

Db 896 NLSDFIA 902

RESULT 9

АДП84078

AAU84078
ID : AAU84078 standard; peptide; 902 AA.XX
IT
AA064076 E

AA784078:

[illegible]

DT 08-MAY-2002 (first entry)

XX
 10-11-68
 (Page 2)

DE S. pneumoniae derived chimeric peptide, VP121.

XX

KW BVH-3; BVH-11; vaccine; mo

KW pneumonia; strep

KW BVH-11-2.

XX

Streptococcus pneumoniae.

OS Synthetic.

XX
09300109334-37

PN -
YY
WO200198334-A2.

27-DEC-2001.
19-JUN-2001; 2001WO-CA00908.
20-JUN-2000; 2000US-212683P.
(SHIR-) SHIRE BIOCHEM INC.
Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
WPI; 2002-132272/16.
New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
epitope-bearing polypeptides, useful as vaccine components for treating
or preventing streptococcal infections such as otitis media,
meningitis, and bacteraemia
Claim 1; Page -: 113pp; English.

PS Claim 1; Page -; 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g. caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
CC

CC pneumococcal genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

AA	Sequence	902 AA;
80		

Query Match	98.4%	Score	4652.5	DB	23	Length	902
Best Local Similarity	99.0%	Pred. NO.	2e-270				
Matches	898	Conservative	0	Mismatches	2	Indels	7
						Gaps	2

QY 1 MQITYDDEIQVAKLAGKYTTEDGYIFDTSWIKKDSLSEAAAAQYAKEKGLTPPSTD 60

2 MOITYDDEIOQVAKLAGKYTTEGYIFDTSWIKKDSIEAERAAQAAYAKEGLTPPSTD 61

QY 61 HQDSGNTAEAKGAEAIYNRVKAAKKVPLDRMPYNLQYTVEVKNGLIIPSYDHYHNIKFEW 120

Db	62	HDSGGNTAEKAGAEIYNRVKAARKVPLDRPNTNLOQTYEVKNGSLIIPSYDYNHNIKFEW	121
QY	121	FDGLYEAPKGYSLIEDLLATVYIYVEPRNASDVRKRNKADQDSKDEKHEVSEPTHP	180
Db	122	FDGLYEAPKGYSLIEDLLATVYIYVEPRNASDVRKRNKADQDSKDEKHEVSEPTHP	181
QY	181	ESDEKENHAGLNPSADNLKPSDTHETEETEEAEADTDDEAETPGTSPISQNAWETLGLKS	240
Db	182	ESDEKENHAGLNPSADNLKPSDTHETEETEEAEADTDDEAETPGTSPISQNAWETLGLKS	241
QY	241	SLLLGTKDNNTISAEVDSYLLALLKESQAPIQ-GPQIGQPTLPNNSLATPSPSLPINPGT	299
Db	242	SLLLGTKDNNTISAEVDSYLLALLKESQAPIQSGPQIGQPTLPNNSLATPSPSLPINPGT	301
QY	300	SHEKHEBDGYGFDPANRIIADESGFVNSHGDSNHVFFKKDLTEBOQIAAKRHLEVKYTS	359
Db	302	SHEKHEBDGYGFDPANRIIADESGFYNS-----YFFKKDLTEBOQIAAKRHLEVKYTS	355
QY	360	NGLDSLSHSHQDYPGNAKEMKDLKKIEKIAGIMKOYGVKRESIVWAKRNAIITYPSGD	419
Db	356	NGLDSLSHSHQDYPGNAKEMKDLKKIEKIAGIMKOYGVKRESIVWAKRNAIITYPSGD	415
QY	420	HHHADPTIDEHKPVGIGHSHSNYELFKPEBGAKEGNKYVTGEELTNVYNLKNSTFNNQ	479
Db	416	HHHADPTIDEHKPVGIGHSHSNYELFKPEBGAKEGNKYVTGEELTNVYNLKNSTFNNQ	475
QY	480	NFTLANGOKRVSPFPPELEKKLGINMLVKLITPDGKVLKESGKGVGGVGNIANFELD	539
Db	476	NFTLANGOKRVSPFPPELEKKLGINMLVKLITPDGKVLKESGKGVGGVGNIANFELD	535
QY	540	QPYLPGQTFKYTTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQA	599
Db	536	QPYLPGQTFKYTTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQA	595
QY	600	VPKGTDALVRVDFEFGNAYLENNYKVGEIKLIPKLNQGTTRTAGNKIPVTTFMANAYLD	659
Db	596	VPKGTDALVRVDFEFGNAYLENNYKVGEIKLIPKLNQGTTRTAGNKIPVTTFMANAYLD	655
QY	660	NQSYIYVEVPTLEKENQTDKPSILPQPKRNKAQENSKLDEKVEEPTSEKVEKEKLSG	719
Db	656	NQSYIYVEVPTLEKENQTDKPSILPQPKRNKAQENSKLDEKVEEPTSEKVEKEKLSG	715
QY	720	NSTNSNLTLEEVPYDVPQVKVAKFAESYGMKLENVLFNMDGTTIELYLPGEVIKKNAADF	779
Db	716	NSTNSNLTLEEVPYDVPQVKVAKFAESYGMKLENVLFNMDGTTIELYLPGEVIKKNAADF	775
QY	760	TGEAPOGNGENKPSNGKVSQGTGVENOPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPE	839
Db	776	TGEAPOGNGENKPSNGKVSQGTGVENOPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPE	835
QY	840	GNVSGDPMPLDPALEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGVEIKK	899
Db	836	GNVSGDPMPLDPALEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGVEIKK	895
QY	900	NLSDFIA 906	
Db	896	NLSDFIA 902	
RESULT 10			
AAU84069			
ID	AAU84069 standard; Peptide; 895 AA.		
XX			
AC	AAU84069;		
XX			
DT	-08-MAY-2002 (first entry)		
XX			
DE	S. pneumoniae derived chimeric peptide, VP111.		
KW	BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;		
KW	pneumonia; streptococcal bacterial infection; mutant; muteln;		
KW	BVH-11-2.		

-ptococcus pneumoniae.
synthetic.

AA PN WO200198334-A2.

XX PD 27-DEC-2001.

XX XX 19-JUN-2001; 2001WO-CA00908.

XX XX 20-JUN-2000; 2000US-212683P.

XX XX (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX DR WPI; 2003-122272/15.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

XX PS Claim 1; Page -; 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Scaphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX SQ Sequence 895 AA;

Query Match 97.8%; Score 4624; DB 23; Length 895;
Best Local Similarity 98.7%; Pred. No. 1e-268;
Matches 894; Conservative 0; Mismatches 0; Indels 12; Gaps 2;

QY 1 MQITYTDEIQVAKLAGYTTEDGYIFDTSNWKDSLSAEARAAQAYAKEGLTPPSTD 60
DB 2 MQITYTDEIQVAKLAGYTTEDGYIFDTSNWKDSLSAEARAAQAYAKEGLTPPSTD 61
QY 61 HODSGNTEAKGAEIYNVRAAKVPLDRPYNLYQYVEVKNGLIIPSDYHYNKFEW 120
DB 62 HODSGNTEAKGAEIYNVRAAKVPLDRPYNLYQYVEVKNGLIIPSDYHYNKFEW 121
QY 121 FDEGLYAPRGYSLEDLATVYKYVEPRNADSHVRKNKADQSDKPDDEKHEVSEPTHP 180
DB 122 FDEGLYAPRGYSLEDLATVYKYVEPRNADSHVRKNKADQSDKPDDEKHEVSEPTHP 181
QY 181 ESEKENHAGLNPSADNLYKPSDTTETEEAEEDTTDEAEIPGTPSRQNAETLTGLKS 240
DB 182 ESEKENHAGLNPSADNLYKPSDTTETEEAEEDTTDEAEIPGTPSRQNAETLTGLKS 241
QY 241 SLLGTGKNDNTISAEVDSLLALKESQAPIQGPQIQGPQIPQNPNSLATPSPSPINPGTS 300
DB 242 SLLGTGKNDNTISAEVDSLLALKESQAPIQGPQIQGPQIPQNPNSLATPSPSPINPGTS 301

QY 301 HEKHEDEGYGFDANRIIADESGFVMSHGSDSNHYFFPKDLTTEQIKAAOKHLEEVKTSN 360
DB 302 HEKHEDEGYGFDANRIIADESGFVMS-----YFFKDLTTEQIKAAOKHLEEVKTSN 355
QY 361 GLDLSHSEQDYPGNAKEMKDLDKKIEEKIAGIMQYGVKRESIVVNNKKNALIIYPSGDH 420
DB 356 GLDLSHSEQDYPGNAKEMKDLDKKIEEKIAGIMQYGVKRESIVVNNKKNALIIY- 411
QY 421 HHADPTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVNLKKNSTNNQN 480
DB 412 --ADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVNLKKNSTNNQN 469
QY 481 FTLANGQKRVSFSPFPELEKKGILNMLVKLITPDGKLVLEKVGKVFEGVGNIANFELDQ 540
DB 470 FTLANGQKRVSFSPFPELEKKGILNMLVKLITPDGKLVLEKVGKVFEGVGNIANFELDQ 539
QY 541 PYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYFPHAGDTYLRVNPQFV 600
DB 530 PYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMASQTFYFPHAGDTYLRVNPQFV 589
QY 601 PKGTDALYRVDFEHGNAYLENNYKVEIKLPIPKLNOGTTTAGNKIPVTFMANAYLDN 660
DB 590 PKGTDALYRVDFEHGNAYLENNYKVEIKLPIPKLNOGTTTAGNKIPVTFMANAYLDN 649
QY 661 QSTYIYEVPILEKENOTDKPSILPQFKRKAQENSKLDEKVEEPKTSKVEKEKLSGTGN 720
DB 650 QSTYIYEVPILEKENOTDKPSILPQFKRKAQENSKLDEKVEEPKTSKVEKEKLSGTGN 709
QY 721 STNSSTLEEYPTVDPQVQKVAFAESYGMKLENVLFNMOGTIELYLPSEGVYIKKNMAFT 780
DB 710 STNSSTLEEYPTVDPQVQKVAFAESYGMKLENVLFNMOGTIELYLPSEGVYIKKNMAFT 769
QY 781 GEAPQNGENKPSKNGKVSQVSTVENOPTENKPADSLPEAPNEKPVKPESTNDGMLNPEG 840
DB 770 GEAPQNGENKPSKNGKVSQVSTVENOPTENKPADSLPEAPNEKPVKPESTNDGMLNPEG 839
QY 841 NVGSDPMLDPALEAPAVDPVQKLEKFTASYGLDSDVIFNMDGTIELRLSPSEGVYIKKN 900
DB 830 NVGSDPMLDPALEAPAVDPVQKLEKFTASYGLDSDVIFNMDGTIELRLSPSEGVYIKKN 889
QY 901 LSDFIA 906
DB 890 LSDFIA 895
RESULT 11
AA084070
ID AA084070 standard; Peptide; 895 AA.
XX AC AA084070;
XX AC
DT 08-MAY-2002 (first entry)
XX S. pneumoniae derived chimeric peptide, vp112.
DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX OS Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
PN 27-DEC-2001.
XX 19-JUN-2001; 2001WO-CA00908.
XX 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 XX epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 PS Claim 1; Page -: 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX
 SQ Sequence 895 AA;

Db 470 FTLANGQKRVSFSPPELEKLGINMLVKLITPDGKLVKSGVFGVGVGNIAFELQ 529
 Qy 541 PVLPGQTFKTIASKDYPVSDGTFTVPTSLAYKMASQTIFFPHAGDTVLRVNPQFV 600
 Db 530 PVLPGQTFKTIASKDYPVSDGTFTVPTSLAYKMASQTIFFPHAGDTVLRVNPQFV 589
 Qy 601 PKGTDALVRVDFEFGHAYLNNKVGKIPKLNQGTTRTAGNKIPVTFMANAYLDN 660
 Db 590 PKGTDALVRVDFEFGHAYLNNKVGKIPKLNQGTTRTAGNKIPVTFMANAYLDN 649
 Qy 661 QSTYIVVPILEKENQTDKPSILPOFRKKAQENSKLDEKVEEPTSEKVEKLSGTGN 720
 Db 650 QSTYIVVPILEKENQTDKPSILPOFRKKAQENSKLDEKVEEPTSEKVEKLSGTGN 709
 Qy 721 STNSSTLEEVTPDVPQVKAFKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNADFT 780
 Db 710 STNSSTLEEVTPDVPQVKAFKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNADFT 769
 Qy 781 GEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEG 840
 Db 770 GEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEG 829
 Qy 841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSVIFNMDGTIELRLPSEGVYKKN 900
 Db 830 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSVIFNMDGTIELRLPSEGVYKKN 889
 Qy 901 LSDFTA 906
 Db 890 LSDFTA 895
 RESULT 12
 AA084073
 ID AA084073 standard; Peptide; 895 AA.
 AC AA084073;
 DT 08-MAY-2002 (first entry)
 XX S. pneumoniae derived chimeric peptide, VP115.
 DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein;
 KW BVH-11-2.
 XX Streptococcus pneumoniae.
 OS Synthetic.
 OS WO200198334-A2.
 PN 27-DEC-2001.
 PD 19-JUN-2001; 2001WO-CA00908.
 PF 20-JUN-2000; 2000US-212683P.
 PR (SHIR-) SHIRE BIOCHEM INC.
 PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 PS Claim 1; Page -: 113pp; English.
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

(I) is useful for therapeutic or prophylactic treatment of otitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.

Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Query Match 97.68; Score 4615; DB 23; Length 895;
 Best Local Similarity 96.68; Pred. No. 3.56-268;
 Matches 893; Conservative 0; Mismatches 1; Indels 12; Gaps 2;

1 MQIYTTDEIQVAKLAGKTYTDEGTYIDTSWIKKDSLEAERAAQAYAKEKGLTPPSTD 60
 2 MQIYTTDEIQVAKLAGKTYTDEGTYIDTSWIKKDSLEAERAAQAYAKEKGLTPPSTD 61
 61 HODSGNTEAKGAEIYNRVRAAKVPLDRMPYNYQYTVVKNGLSLIIPSDYHYNIKFEW 120
 62 HODSGNTEAKGAEIYNRVRAAKVPLDRMPYNYQYTVVKNGLSLIIPSDYHYNIKFEW 121
 121 FDEGLYAPKGYSLIEDLATVYKYYVEPRNADSHVRKKAQDSDPKDEHDESETHP 180
 122 FDEGLYAPKGYSLIEDLATVYKYYVEPRNADSHVRKKAQDSDPKDEHDESETHP 181
 181 ESKENHAGLNPSADNLYKPTDTEETEEAEDTTDEAETPGTFSIRQNAETLTGLKS 240
 182 ESKENHAGLNPSADNLYKPTDTEETEEAEDTTDEAETPGTFSIRQNAETLTGLKS 241
 241 SLLGTGDNNTISAEVDSLLALKESOPAPITQGPQIGPILPNNSLATPSPSLINFGTS 300
 242 SLLGTGDNNTISAEVDSLLALKESOPAPITQGPQIGPILPNNSLATPSPSLINFGTS 301
 301 HEKHEEDGDFDANRIITAEDSGFVMSHGSDNHYFFKDLTEEOIKAAQKLEEVKTSN 360
 302 HEKHEEDGDFDANRIITAEDSGFVMS-----YFFKDLTEEOIKAAQKLEEVKTSN 355
 361 GLDSSLSEHQDYPGNAKEMKDLKKIEKIAGIMKQYGVKRESIVVAKENAIYPSGDH 420
 356 GLDSSLSEHQDYPGNAKEMKDLKKIEKIAGIMKQYGVKRESIVVAKENAIYYP--- 411
 421 HHADPIDEKHPVGIHSHSNYELFKPEGVAKKGNKYVYTGEEELTNVLLKNSTFNNQN 480
 412 --ADPIDEKHPVGIHSHSNYELFKPEGVAKKGNKYVYTGEEELTNVLLKNSTFNNQN 469
 481 FTLANGQRVVSFPPELEKLGINMLVKLITPDGKLVKESGVFGVGVGNIAFELDQ 540
 470 FTLANGQRVVSFPPELEKLGINMLVKLITPDGKLVKESGVFGVGVGNIAFELDQ 529
 541 PYZLPGQFKTYTISKDPEVSYDGTFTVPTSLAYKMASQTIFFPHFAGDTYLRVNPQFV 600
 530 PYZLPGQFKTYTISKDPEVSYDGTFTVPTSLAYKMASQTIFFPHFAGDTYLRVNPQFV 589
 601 PKGTDALVRVDFEPHGNAYLNNYKVEIKLPIKLNQGTTRTAGNKIPTFFMANAYLDN 660
 590 PKGTDALVRVDFEPHGNAYLNNYKVEIKLPIKLNQGTTRTAGNKIPTFFMANAYLDN 649
 661 QSTYIVVEVPILEKENQDKPSILPQFRNKAQENSKLDEKVEEPTSEKVEKELSETGN 720
 650 QSTYIVVEVPILEKENQDKPSILPQFRNKAQENSKLDEKVEEPTSEKVEKELSETGN 709

QY 721 STNSTLEEVPTDPOEVKAKFAESYGMKLENVLPNMDGTIELYLPSPGVEIKKNMADFT 780
 DB 710 STNSTLEEVPTDPOEVKAKFAESYGMKLENVLPNMDGTIELYLPSPGVEIKKNMADFT 769
 QY 781 GEAPQNGENKPSGKSTGTSTVENQPTENKPADSLPEAPNEKPKVPKPNSTONGMLNPEG 840
 DB 770 GEAPQNGENKPSGKSTGTSTVENQPTENKPADSLPEAPNEKPKVPKPNSTONGMLNPEG 829
 QY 841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVFNMDGTIELRLPSPGVEIKKN 900
 DB 830 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVFNMDGTIELRLPSPGVEIKKN 889
 QY 901 LSDFIA 906
 DB 890 LSDFIA 895

RESULT 13
 AAU84075
 ID AAU84075 standard; Peptide; 895 AA.
 XX
 AC AAU84075;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE S. pneumoniae derived chimeric peptide, VP117.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mtein;
 KW BVH-11-2.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 XX
 PN W0200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA00908.
 XX
 PR 20-JUN-2000; 2000US-212683P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX
 DR WPI; 2002-122272/16.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 PS Claim 1; Page -: 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

QY	Sequence	895 AA;
	Query Match	97.69; Score 4615; DB 23; Length 895;
	Best Local Similarity	98.69; Pred. No. 3.5e-268;
	Matches 893; Conservative	0; Mismatches 1; Indels 12; Gaps 2;
QY	1 MQIYTTDEIQVAKLAGKTYTDEGYIFDTSHIKKDSLSEARAAQAAYAKEKGLTPPSTD	60
DB	2 MQIYTTDEIQVAKLAGKTYTDEGYIFDTSHIKKDSLSEARAAQAAYAKEKGLTPPSTD	61
QY	61 HQDSGNTAEKAGAEIYNNRVKAAKVPDLDRMPYNLQYTYEVANGSLIIPSYDHYHNKFEW	120
DB	62 HQDSGNTAEKAGAEIYNNRVKAAKVPDLDRMPYNLQYTYEVANGSLIIPSYDHYHNKFEW	121
QY	121 FDEGLYEAPKGYSLEDDLATVKKYVYVEPNASHVKNKADQSKDPDEKHEHDEVSEPTHP	180
DB	122 FDEGLYEAPKGYSLEDDLATVKKYVYVEPNASHVKNKADQSKDPDEKHEHDEVSEPTHP	181
QY	181 ESDKEKHAGLNPSADNLYKPSDTETEESAETDTDEAEIPGTPSIRONAMETITGLKS	240
DB	182 ESDKEKHAGLNPSADNLYKPSDTETEESAETDTDEAEIPGTPSIRONAMETITGLKS	241
QY	241 SLLIGTRDNNTISAEVDSLLALLKESAPATQGPQIGOPTLPNNSLATPSPSLPINPCTS	300
DB	242 SLLIGTRDNNTISAEVDSLLALLKESAPATQGPQIGOPTLPNNSLATPSPSLPINPCTS	301
QY	301 HEKHEEDGYGPDANRIIAEDSGFVMSHGDSNHYFFFKDLTEEQIKAAQKHLVEYKTSN	360
DB	302 HEKHEEDGYGPDANRIIAEDSGFVMS-----YFFKKDLTEEQIKAAQKHLVEYKTSN	355
QY	361 GLDSLSSHEQDYPGNAMKMDLDKIEKIAGIMKQYGVKRESIVVNEKKNAIITPSSDH	420
DB	356 GLDSLSSHEQDYPGNAMKMDLDKIEKIAGIMKQYGVKRESIVVNEKKNAIITP-----	411
QY	421 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKGNKYVTGEELTNVYNLAKNSTFNNQN	480
DB	412 --ADPIDEHKPVGIGHSHSNYELFKPEEGVAKKGNKYVTGEELTNVYNLAKNSTFNNQN	469
QY	481 FTLANGKRVSFSPPELEKKLQINLVLKLTDPCKVLKESGVKFGVGGVGNIAINFELQ	540
DB	470 FTLANGKRVSFSPPELEKKLQINLVLKLTDPCKVLKESGVKFGVGGVGNIAINFELQ	529
QY	541 PYPGQTFKTYIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAY	600
DB	530 PYPGQTFKTYIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAY	589
QY	601 PKGPDALVRVDEPHGNAYLNNYKVGSEIKLPIPKLNGQTTTTRAGNKIPVTFMANAYLDN	660
DB	590 PKGPDALVRVDEPHGNAYLNNYKVGSEIKLPIPKLNGQTTTTRAGNKIPVTFMANAYLDN	649
QY	661 QSTYIVVEVPILEKENQTDKPSILPOPKRKAQENSKLDEKVEEPTSKVKEKELSETGN	720
DB	650 QSTYIVVEVPILEKENQTDKPSILPOPKRKAQENSKLDEKVEEPTSKVKEKELSETGN	709
QY	721 STNSSTLEEYPTVPQVEKVAFAESYGMKLENVLNMDGTYIELYLPSEGVYKKNMADFT	780
DB	710 STNSSTLEEYPTVPQVEKVAFAESYGMKLENVLNMDGTYIELYLPSEGVYKKNMADFT	769
QY	781 GEAPQNGENKPSNGKYSTGTGVNQPTENKPADSLPEAPNEKPKVPKENSTDNGMLNPEG	840
DB	770 GEAPQNGENKPSNGKYSTGTGVNQPTENKPADSLPEAPNEKPKVPKENSTDNGMLNPEG	829
QY	841 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSPSGEVIKN	900
DB	830 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSPSGEVIKN	889
QY	901 LSPDIA	906

BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia; streptococcal bacterial infection; mutant; muten;	BVH-11-2.
Streptococcus pneumoniae.	Synthetic.
WO200198334-A2.	
27-DEC-2001.	
19-JUN-2001; 2001WO-CA00908.	
20-JUN-2000; 2000US-212683P.	
(SHIR-) SHIRE BIOCHEM INC.	
Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;	
WPI; 2002-122272/16.	
New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia	
Claim 1; Page -; 113pp; English.	
The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.	
Note: This sequence does not appear in the specification but has been created according to information given in the invention.	
Sequence 896 AA;	
Query Match 97.5%; Score 4608.5; DB 23; Length 896;	
Best Local Similarity 98.5%; Pred. No. 8.7e-266;	
Matches 893; Conservative 0; Mismatches 1; Indels 13; Gaps	
QY 1 MQITYTDDETOAVKAGKYYTDEGYIFDTSWIKKDSLSEAEAAQAAYAEKGLTPPSTD 60	
DB 2 MQITYTDDETOAVKAGKYYTDEGYIFDTSWIKKDSLSEAEAAQAAYAEKGLTPPSTD 61	
QY 61 HQDSGNTKAGAEIYNNRYKAAKYPLODMYPNLYQTYVEVNGSLIIPSYDHYHNKFEW 130	
DB 62 HQDSGNTKAGAEIYNNRYKAAKYPLODMYPNLYQTYVEVNGSLIIPSYDHYHNKFEW 115	
QY 121 FDEGLYEPKGYSLIEDLLATVYTYVEPRNASDHYRKNKADQDSPDEKHEDEVSETHP 180	
DB 116 FDEGLYEPKGYSLIEDLLATVYTYVEPRNASDHYRKNKADQDSPDEKHEDEVSETHP 175	
QY 181 ESDKEHNAGLNPSADNLKYPSTDTETEEDTDEAEIPGTPPSIRONAMEYLTGLKS 240	
DB 176 ESDKEHNAGLNPSADNLKYPSTDTETEEDTDEAEIPGTPPSIRONAMEYLTGLKS 235	

QY 241 SLLGKDNNTISAEVDSLALLKESQAPAIQ-GPQIGOPTLPNNSLATPSPSLPINPCT 299
 Db 236 SLLGKDNNTISAEVDSLALLKESQAPAIQSGPQIGOPTLPNNSLATPSPSLPINPCT 295
 QY 300 SHEKHEEDGYGDANRIIAEDSGFVMSHSDSNHFFKDLTEBQIKAAQKHLEEVKTSH 359
 Db 296 SHEKHEEDGYGDANRIIAEDSGFVMS-----YFFKDLTEBQIKAAQKHLEEVKTSH 349
 QY 360 NGLDSLSSHEQDYPGNAKEMKDLKIEKIAKIMQGVKRESIVVWNEKNAIYPSGD 419
 Db 350 NGLDSLSSHEQDYPGNAKEMKDLKIEKIAKIMQGVKRESIVVWNEKNAIYPSGD 409
 QY 420 HHADPIDEHKPVGICHSHSNTYELFKPEGVAKKEGKNYITGEELTNVNVNLLKNSTFNNQ 479
 Db 410 HHADPIDEHKPVGICHSHSNTYELFKPEGVAKKEGKNYITGEELTNVNVNLLKNSTFNNQ 469
 QY 480 NFTLANGOKRVSFSPPELEKLGIMLVKLITPDGKLVLEKYSKVFGEVGNIANFELD 539
 Db 470 NFTLANGOKRVSFSPPELEKLGIMLVKLITPDGKLVLEKYSKVFGEVGNIANFELD 529
 QY 540 OYPLPGQTFKYTIASKDYPEVSDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFA 599
 Db 530 OYPLPGQTFKYTIASKDYPEVSDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFA 589
 QY 600 VPKGTDALVRVDFEFGHGNAYLENNYKVGGEIKLPIPKLNGQTTTAGNKIPVTFMANAYLD 659
 Db 590 VPKGTDALVRVDFEFGHGNAYLENNYKVGGEIKLPIPKLNGQTTTAGNKIPVTFMANAYLD 649
 QY 660 NOSTYIVVEPILKEKENTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSG 719
 Db 650 NOSTYIVVEPILKEKENTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSG 709
 QY 720 NSTNSTLEEVPDVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKKNMADF 779
 Db 710 NSTNSTLEEVPDVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKKNMADF 769
 QY 780 TGEAPQNGENKPSNGKVSSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPE 839
 Db 770 TGEAPQNGENKPSNGKVSSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPE 829
 QY 840 GNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSEGEVKK 899
 Db 830 GNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSEGEVKK 889
 QY 900 NLSDFIA 906
 Db 890 NLSDFIA 896

Search completed: March 27, 2003, 11:54:24
 Job time : 47 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 11:53:37 ; Search time 20 Seconds
(without alignments)
1332.858 Million cell updates/sec

Title: US-09-884-465A-332

Perfect score: 4728

Sequence: 1 MQITYTDEIQVAKLAGKYT.....TELRLPSGEVKNLSDFIA 906

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	818.5	17.3	763	4	US-08-961-083-56
2	636	13.5	796	4	US-08-961-083-56
3	431	9.1	447	4	US-08-961-083-182
4	233.5	4.7	10182	4	US-08-961-083-3159
5	199.5	4.2	1098	4	US-08-923-992A-8
6	195.5	4.1	1128	4	US-08-923-992A-6
7	194.5	4.1	1164	4	US-08-923-992A-2
8	192	4.1	571	4	US-08-961-083-4
9	192	4.1	1164	4	US-08-923-992A-10
10	191.5	4.1	3696	4	US-08-923-992A-5080
11	185.5	3.9	1104	4	US-08-923-992A-8
12	184.5	3.9	1964	2	US-08-790-912-3
13	184.5	3.9	2052	2	US-08-790-912-2
14	182	3.8	984	1	US-08-242-932-2
15	182	3.8	984	1	US-08-714-481-2
16	182	3.8	984	5	PCT-US95-06111-2
17	180.5	3.8	1183	2	US-08-447-031A-2
18	175.5	3.7	1231	3	US-08-904-263A-4
19	175	3.7	1183	4	US-08-134-001C-3530
20	170.5	3.6	1545	4	US-08-286-791-4
21	170.5	3.6	1545	5	PCT-US95-10661A-4
22	170	3.6	1848	4	US-08-236-791-6
23	170	3.6	1848	5	PCT-US95-10661A-6
24	168.5	3.6	3135	1	US-08-323-170B-2
25	168.5	3.6	3135	4	US-08-954-441-2
26	167	3.5	783	6	5231168-2
27	165.5	3.5	1181	4	US-09-206-898-23

28	164.5	3.5	962	4	US-09-071-035-246	Sequence 246, App
29	164.5	3.5	962	4	US-09-071-035-250	Sequence 250, App
30	164.5	3.5	962	4	US-09-071-035-254	Sequence 254, App
31	164.5	3.5	962	4	US-09-071-035-470	Sequence 470, App
32	164.5	3.5	962	4	US-09-071-035-474	Sequence 474, App
33	164.5	3.5	962	4	US-09-071-035-478	Sequence 478, App
34	164	3.5	1164	2	US-08-589-756-1	Sequence 1, Appl1
35	164	3.5	1164	4	US-09-206-800-1	Sequence 1, Appl1
36	164	3.5	1164	4	US-09-206-898-1	Sequence 1, Appl1
37	163.5	3.5	1881	4	US-09-233-086-3	Sequence 3, Appl1
38	162	3.4	2123	4	US-08-988-885A-10	Sequence 10, Appl1
39	161	3.4	676	4	US-09-134-001C-4318	Sequence 4318, Ap
40	161	3.4	1167	2	US-08-589-756-2	Sequence 2, Appl1
41	161	3.4	1167	4	US-09-206-800-2	Sequence 2, Appl1
42	161	3.4	1167	4	US-09-206-898-2	Sequence 2, Appl1
43	160.5	3.4	1301	4	US-09-071-035-234	Sequence 234, App
44	160.5	3.4	1301	4	US-09-071-035-238	Sequence 238, App
45	160.5	3.4	1301	4	US-09-071-035-242	Sequence 242, App

ALIGNMENTS

RESULT 1
US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 5231168
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: FB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8504
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-66

Query Match 17.3% Score 818.5; DB 4; Length 763;
Best Local Similarity 66.1% Pred. No. 4.4e-52;
Matches 168; Conservative 16; Mismatches 29; Indels 41; Gaps 5;
Qy 2 QITYTDEIQVAKLAGKYTTEDGIFD-----TSWIKKDSLSSEAR 42
Db 498 QITYTDEIQVAKLAGKYTTEDGIFDPRDITSDEGDYVTPHMTSHWIKKDSLSSEAR 557

QY 43 AAQAYAKEGLTPPSTDHQDSGNTGAEAGAIYNRVKAARKVPLDRMPYNLQYTVYVK 102
 DB 558 AAQAYAKEGLTPPSTDHQDSGNTGAEAGAIYNRVKAARKVPLDRMPYNLQYTVYVK 617
 QY 103 GSLIIPSDYHNIKFDFGLYEAPKGYSLDGLLATVYKYYE-----PR-----NAS 151
 DB 618 GSLIIPSDYHNIKFDFGLYEAPKGYSLDGLLATVYKYYE-----PR-----NAS 151
 QY 152 DHVRKNNK-----ADQSKPDKDEHDEVPSETHPESDEKENHAGLNPSADNLYKPSDTTE 206
 DB 678 DHVRKNNKQADTNOTKPSSEKPTKPEETPREKPSQSEKPEP-----RPTPEE 731
 QY 207 ETEEAEDTDEAE 220
 DB 732 ESPESEEPQVETE 745

RESULT 2
 US-08-961-083-56
 ; Sequence 56, Application US/08961083
 ; Patent No. 6159469
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/961,083
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 56:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 796 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-961-083-56

Query Match 13.58; Score 636; DB 4; Length 796;
 Best Local Similarity 45.38; Pred. No. 1,46-38;
 Matches 140; Conservative 40; Mismatches 75; Indels 54; Gaps 8;
 QY 2 QITTDDEIQVAKLAGKTTTDEGIFD-----TSWIKKDSLSSEAR 42
 DB 507 QIETDEVRIAQLADKYTSDGYTFDDEHDIISDEGDAYVTPFHGSHWICKSDLSDEK 566
 QY 43 AAQAYAKEGLTPPSTDHQDSGNTGAEAGAIYNRVKAARKVPLDRMPYNLQYTVYVK 102
 DB 567 VAAQATTEKGLPPSPDQVANKPTGDSAAAIYNRVKEKRIPVLRPLVNVHEHTVEYKN 626

QY 103 GSLIIPSDYHNIKFDFGLYEAPKGYSLDGLLATVYKYYE-----PR-----NAS 151
 DB 627 GNLIIPKDHVHNIKFDFGLYEAPKGYSLDGLLATVYKYYE-----PR-----NAS 151
 QY 152 DHVRKNNKADQSKPDKDEHDEVPSETHPESDEKENHAGLNPSADNLYKPSDTTEETEE 211
 DB 687 EHV-LGKDKHSEDPKKNKFADE--EPVEETPAEPE-----VPOVETKVEAQ 730
 QY 212 AEDTTDAEAI-----PGTPSIHQNAWETLTGLKSSLLGLTKDNTTISAFVDSLALLAKESQ 267
 DB 731 LK-----EAEVLAKVTOSSSLKANATETLAGLUNNLTLOIMDNNSIMAEAEKLLALLAGSN 786
 QY 268 PAPIQGPQI 276
 DB 787 PSSVSKEKI 795

RESULT 3
 US-08-961-083-182
 ; Sequence 182, Application US/08961083
 ; Patent No. 6159469
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/961,083
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 182:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 447 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-961-083-182

Query Match 9.18; Score 431; DB 4; Length 447;
 Best Local Similarity 32.28; Pred. No. 7,7e-28;
 Matches 123; Conservative 40; Mismatches 91; Indels 128; Gaps 12;
 QY 7 DDEIQVAKLAGKTTTDEGIFD-----DTS-----WIKKDSLSSEARAAQA 47
 DB 144 NSNVAVARSQRYTNDGYTFVFNPAIIETGNAVIVPHGGHYHYPKSDLSASELAAGA 203
 QY 48 YAKEKGLTP-----PSTDHQSNGTAEAGA-----EAIYNRVKAARKVPL 87
 DB 204 HLAKNNQPSQLSYSTASDNNTQSVAGSTSPANKSENLSLKLKDYSPSAORYSES 263
 QY 88 DRMPYNLQYTVYKNGSLIIPSYDHYHNKFNWFEGLYEAPKGYSLDGLLATVYKYYE 147

APPLICATION NUMBER: US 60/024,707
FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1098 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-923-992A-8

Query Match 4.28; Score 199.5; DB 4; Length 1098;
Best Local Similarity 17.68; Pred. No. 3.6e-06;
Matches 157; Conservative 125; Mismatches 289; Indels 321; Gaps 34;

QY 166 DEDKEHDEVSEPTHP---ESDEKENHAGLNPSADNLYKPSDTTETEEAEADTTDEAIP 222
DB 12 DSVKTEVAAPKPPSMAQTDOGNSSSELETRMEIPTDIKKAPEVPEKTAGETSAT 71
QY 223 GTPSIRONAMETLGLKSSL---LLGTNDNTISAEV-----DSLAL----- 262
DB 72 HTGKREKLOQWKNLNDVDNTILSHQKNEFKTKIDETNDSALLEENQFNETNRL 131
QY 263 -LKESQAPAPGPOIGOPTLNNSLATSPSLPNPGTSHKHEEDGYGFANDRIIAEDE 321
DB 132 HIKHEEVEKDKAKAQQTKQSD--TVVDLSNIDKELNHQSKQENG-----ITNEDK 182
QY 322 SGFVMSHGDNNHFFPKDLTE-----FOIK----- 346
DB 183 DSMKAKIEDIRKQAQPKDKEAEVKKVEELGKLFSSTKAGLDQEIQSHVVKETSSE 242
QY 347 -----AAKHLEEVKTSNGLDLSLSEHQ-----DYPGNKAKEMKDLKKIE 387
DB 243 QKVDEHYANSQNLAKSLEE-----LDKATNEQATQVKNQFLENQAQKLEIPLIK 295
QY 388 EKIAGIKQYGVKRESI-----VVKENKNAIYPSGDHHPIDIEHKPVG 433
DB 296 ETNVLKRAMSSLSQVEKELKHSEANLELDVAKSKEIVREYEGKLNQSKNLPKQLE 355
QY 434 ICHSHNTELFEPGVAKEGKNTYTBELTNVNLKNSTFNQNTLANGQRVSFS 493
DB 356 -BEAHSKL-----KQVVEDFRKFKTSQVTPKRLKRDLAANE-----NNOOKIELT 402
QY 494 PPELEKKGILNMVKLITPDGKLVKSGKVGEGVGNIANFELDQVLPQOTFKYITIA 553
DB 403 VSFE-----NITVE-----GEDVKFTVT 421
QY 554 SKDYPEVSDGFTTPTSLAYKMASQTIFFPHAGDTYLRVNPQAPVPGTDALYRVDE 613
DB 422 AKSDSKTLD-----FSDLLTKYNSVS-----DRISTN 450
QY 614 FGNAYLENNYKVGKILPIKLNQGTTRT-----AGNKIPVTWANYLDNQSYIVE 667
DB 451 YKNT-----DNHIAEITTKNLNESQVTLKAKDSDGNVKEFT-----T 493
QY 668 VPILKENQTDKPSILPOPKRKAQENSKLDEKVEEPTSEKVEKEKILSETGN----- 720
DB 494 ITVQKEER-----QVPKPEQKDSKTEKVPQEPKSNQKQLQELIKSAQOQLEKL 545
QY 721 STSNSTLEVPV-----DPVQEKVAKF-----AESYGMKL 751
DB 546 EKAIKELAQEIPSNPEYGIQSIWESQKEPIQEAITSFKRIIGDSSSKYITTEHYFNKY 605
QY 752 ENVLPMQDCTIELYPSGVIKKNMADFTGEAPOGENGENKPSKGVST-----GTVENQ 806
DB 606 KSHFMNYQLHAQM-----EMLTRKVVQYNNKYPDNAEIKKIFESDMKRTKEDNYGSLND 660

QY 807 PTE-----NK-----PADSLPEAPNEKPVKPNSTDNGLNPEGNVGDPM 849
DB 661 ALKGYFEKYLTPFNKIKQIVDDFKKVDQDQAPAPI-PENSE-----MD 703
QY 850 PALEEA-----PAVDPVOEKLEKFTASYGLGL-----DSVIFNMD 884
DB 704 QAEKAKIAVSKYMSKVLGDGVHQHLQKKNHKSIVDLFKELEAIKQOTIFDID 755

RESULT 6
US-08-923-992A-6
Sequence 6. Application US/08923992A
Patent No. 6280738
GENERAL INFORMATION:
APPLICANT: Tai, Joseph Y.
APPLICANT: Blake, Milan S.
TITLE OF INVENTION: No. 6280738-iga Fc Binding Forms of the Group B
TITLE OF INVENTION: Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,992A
FILING DATE: 05-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,707
FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-923-992A-6

Query Match 4.18; Score 195.5; DB 4; Length 1128;
Best Local Similarity 17.38; Pred. No. 7.3e-06;
Matches 154; Conservative 130; Mismatches 296; Indels 309; Gaps 34;

QY 166 DEDKEHDEVSEPTHP---ESDEKENHAGLNPSADNLYKPSDTTETEEAEADTTDEAIP 222
DB 7 DSVKTEVAAPKPPSMAQTDOGNSSSELETRMEIPTDIKKAPEVPEKTAGETSAT 66
QY 223 GTPSIRONAMETLGLKSSL---LLGTNDNTISAEV-----DSLAL----- 262
DB 67 DTGRERQLOQWKNLNDVDNTILSHQKNEFKTKIDETNDSALLEENQFNETNRL 126
QY 263 -LKESQAPAPGPOIGOPTLNN-----NSLATPSPSLPINFQTSHE----- 302
DB 127 HIKHEEVEKDKAKAQQTKQSDTKVLDLSNIDKELNHQSKPVKMAEPKGTITNEDK 186
QY 303 -----KHEEDGYGFANDRIIAEDSGFVMSHGDN-----HYFFKDLTTEQ----- 344
DB 187 LKTIEDINKQAQADKDEAEVKKVEELGKLFSSTKAGLDQEIQSHVVKETSSE 246

345 QY -----IKAAQHLEVKTSNGLDLSLSHEQ-----DYPNAKEMKDLDRKIEKI 390
Db 247 DEHVANSQNLAKSLEE-----LDKATNEQATQVKNQFLENQAKLQKEMOPLIKETN 299
QY 391 AGINKQYGVKRESI-----VWKEKNAYIPSGDHHHADPIDEHPVGLGH 436
Db 300 VKLYKAMESLEQVEKELHNSANLEMDLAKSEIVREYEGKLNQSNLPELKOLE-EE 358
QY 437 SISNYELKPEGVAKKGNVYTGELTNVNLKNSTFNQNTLANGOKRVSFSPFP 496
Db 359 AHSKL-----KQVEDPRKFKFSEQVTPKRVKRLAANE-----NNQKIELTVSP 406
QY 497 ELEKRLGINMLVLLITPDGKLVKSGVPGEGVGNINFAELDQPLYPGOFKTYTIASKD 556
Db 407 E-----NITVIE-----GEOVKTVTAKS 425
QY 557 YPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFVAVPKGTDALVRVDFPHG 616
Db 426 DSKTTLID-----FSDLLTKYNPSVS-----DRISTNTKT 454
QY 617 NAYLENNYKVGKILPIPKINOGTTRT-----AGNKIPVTFMANAYLDNOSTYIVEVPI 670
Db 455 NY-----DNHKAIEITIKNLKESQVTLKAKDDSGNVVEKTF-----TIV 497
QY 671 LKENQOTDPSILPOFRKNAQENSKLDKV-EEPKTSEKVEKEKLSGTN-----STS 723
Db 498 QKKEE-----QVPTPEOKDSKTEKVPQEPKSDNKQLOELIKSAQOQLEKLEKA 549
QY 724 NSTLEVPV-----DPQEKVAKF-----AESYGMKLENV 754
Db 550 IELMEOPEIPSNPEYGIQKSWESQEKIQAITSFKKIIGDSSKYYTEHYFNKYSK 609
QY 755 LFMNDGTIELYPSGEVIRKKNMADFTGEAPOGNGENKPSKSVT-----GTVENQPT 809
Db 610 FNNYQLHAQM-----EMLTRKVYQYANKYPDNAEIKKIPESDKRTKEDNYGSLNDALK 664
QY 810 -----NK-----PADSLPEAPNEKVPKPNSTDNGLNPEGNVGSDDPMDPAL 852
Db 665 GYPEKYTLFPFNKIQIVDLOKKVEQDQAPFI-PENSE-----MDQAK 707
QY 853 EEA-----PAVDPVQEKLEKFTASYGLGL-----DSVIFNMD 884
Db 708 EKAKIAVSKYMSKVLGVGHQHLQKKNHNSKIVDLFKELEAIKQOTIFDID 756

RESULT 7
US-08-923-992A-2
Sequence 2, Application US/08923992A
Patent No. 6280738
GENERAL INFORMATION:
APPLICANT: Tal, Joseph Y.
APPLICANT: Blake, Milan S.
TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B
TITLE OF INVENTION: Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,992A
FILING DATE: 05-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,707

FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1164 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-923-992A-2

Query Match 4.1%; Score 194.5; DB 4; Length 1164;
Best Local Similarity 19.4%; Pred. No. 9.1e-06;
Matches 166; Conservative 130; Mismatches 316; Indels 245; Gaps 40;

QY 166 DEKHEDEVSEPHTP-----ESDEKENHAGLNPSADNLYKPSDTEETEAEADTTDEAEP 222
Db 43 DQSVKTTVAAPKYPSPMAOTDQGNSSSELETTKMBEIPITDIKKAPEVPEKTAGETSAT 102
QY 223 GTPSIRONAMETLTGLKSSL-----LLGPKDNNTISAEV-----DSLIALLKESQAPAIQ 273
Db 103 DTGREKQLOQWKNLNDVNDWILSHEQNEFKTKIDETNDSALLE-----151
QY 274 PQIGOPTLPNNSLATPSPSLPPIPGTSHKHEEDGYGFDANRIAEDESQFVMSHSDS-- 331
Db 152 -----ENQFNETNRLHLIK---QHEVEKDKKA-KQOKTLKQSDTKVDLSNIDKEL 198
QY 332 NH-----FFPKDLTEQIKAAKHLEVKTSNGLDLSLSHEODYPGNAMKMDLKD 384
Db 199 NHQSOVKMAQEGITFEDKDSMLKIEDIRKQAQADKEDAE-----VKVREELGK 252
QY 385 KIEKIAIGIMK---QYGVKRESIVVYKKNAYIPSGDHHHADPI-----426
Db 253 LFSSTKAGLDQEIQEHVKET---SSEET---QKVDHYANSQNLAKSLEELDKATT 306
QY 427 DEHPVGVGHSHSNYELFKPEGVAKKGNKY---TGEELTNVNLKNSTFNQNTLA 484
Db 307 NEQATQVKNQFLENQAKLKEIOTLIKETNVLKYNKAMESLEQVEKELKHNSANLEDLVA 366
QY 485 NGKRVSPFPPELEKLGINMLVKLITPDGKLVKSGVPGEGVGNI-ANFELDQPYL 543
Db 367 KSKEIYR-----EYEGKL--NQSKNL--PELKQLEEEAHSKQKQVEDFRKKFTSEQVT 417
QY 544 PQGTFKYTIASKD-----TPE--VSYDG---TFTVPTSLAYKMASOTIYFPFHAG 588
Db 418 PKKRVKRLAANENNQOKIELTVSPENITVYEGEDVKFTVTA---KSDSKT---TLDFS 470
QY 589 DTYLRVNPQFVAVPKGTDALVRVDFEFGHNAVLENNYKVGKELKPIPKINOGTTRT----- 643
Db 471 DLLTKYNPSVS-----DRISTNYKNT---DNHKAIEITIKNLKESQVTLKAKD 519
QY 644 -AGNKIPVTFMANAYLDNOSTYIVEVPILEKENOTDKPSILPQFKNKAQENSKLDKRV- 701
Db 520 DSGNVVEKTF-----TIVQKKEK-----QVPTPEOKDSKTEKVP 557
QY 702 EEPKTSKVEKEKLSGTN-----STNSNILEEYPTV-----DPVOE 738
Db 558 QEPKSNDRKQLOELIKSAQOQLEKLEKAIKELMEQEPSNPSNPEYGIQKISINESQKEPQ 617
QY 739 KVAKF-----AESYGMKLENVLFNMDGYTIELYPSGEVIRKKNMADFTGEAPOG 786
Db 618 ATTSFKKIIGDSSSKYYTEHYFNKYSDFMNYQLHAQM-----EMLTRKVYQYNNKYPDN 672
QY 787 NGENKPSKSVST-----GTVENQPT-----NK-----PADSLPEAPNEK 824
Db 673 AEIKKIFESDMKRTKEDNYGSLNDALKGYFEKYFLTPFNKIKQIVDLDKRVKVEDQAP 732
QY 825 VRPENSTDNGLNPEGNVGSDDPMDPALAEA-----PAVDPVQEKLEKFTASYGLG 875

Db 733 I-PENSE-----MDOAKEKAKIAVSKYMSKVLGVDGHOHLQKNNKLIYD 775
QY 876 L-----DSVIFNMD 884
Db 776 LPKELEAIKQOIFDID 792
RESULT 8
US-08-961-083-4
; Sequence 4, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 432
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-4
Query Match 4.1%; Score 192; DB 4; Length 571;
Best Local Similarity 22.2%; Pred.No. 4.8e-06;
Matches 156; Conservative 88; Mismatches 197; Indels 262; Gaps 40;
QY 284 NSLATPSPLPDPGSHKEHEDGIGFDAN-----RIIAEDBSFVMSHSDSNHYFFK 337
Db 45 SSVATPT-----KQKQYDYNVTFNFDHPSTVQAIQEQTPVSSTKPTVEQVVE 92
QY 338 KDLTEEOIKAAQKHLKRVKTSNGLDLSSEHEDQPCGNAXEMKDLKKIEKLAGIMKQY 397
Db 93 KPFTELINPR-----KEKQSSDSQEQLAHE-----ANLETKKEEKIS----- 131
QY 398 GVKRESIVVAKENAIYPSGDHHDAPIDE-----HKPVGIGHSHSNYELFKPPEGVA 451
Db 132 -----PEKTKGV-----NTLNPQDVLSSGQLNKP-----ELLYREETME 165
QY 452 KKEGNVYTGELTNVNLKNSTFNQNTLANGQRVFSFPPELEKLGINKM-LVKL 510
Db 166 -----TKIDFQEEI-----QENPDLAGTGYRV-----ROEGLKGGKVEIVRI 202
QY 511 IT-----PDGKLVKVSQK--VFGE-----GV-----GNTIANPELDQ 540

Db 203 FSVNKEEVSREIYSTTAPSPRIVEKTKKQVIKEQPETGVEHKDVQSGAIVERAI-Q 261
QY 541 PYLPQGFKTIASKDYPEVSYDGTFTVPTSLAYKASQTIFPFHAGDTYLRVNPQ--- 597
Db 262 PELP-----EAVVSDKGEPEVOP-----TLPEAVV-----TDKGET--EVQESPD 300
QY 598 -FAVPKGTDALVRVDFHGNAYLENNYKVGEIK--LPKPL-NQGTTRTAGNKIPVTEM 653
Db 301 TVVSDKGEPEQVAPLPEYKGN-----LEQVKPETPVEKTEQGPKEI--EEVPV--- 347
QY 654 ANAYLDNQSTIYVEVPVILEKENQDKPSILPQKRN---FAQENSKLDEKVEPKSEK- 709
Db 348 -----KPTETPVNNEGTTGTSI--QEAENPVQPAEESTTNSEKV-SPDTSSKN 395
QY 710 -----VEKEKLSGTGNSNSTSLTELEVPTDVPQVKVAKFAESYGMKLE 752
Db 396 TGEVSSNPSTTSVSGSNKPEHDSNENSEKTYEVP-VNP----- 437
QY 753 NVLFNMDGTIELYLPSCGEVINKNMADFTGEAPQNGENKPSSENGKV---STGTVENOPT 809
Db 438 -----NEGTV-----GTSNQETKPVQPAEETQTSNGKIANENTGEVSNKPSD 481
QY 810 NKP--ADSLPEAPNEKPVKPPENS-----TDNGLNPE---GNVGSDDPMLDPALEAPAVDP 860
Db 482 SKPPVEESNQPEKNGTATKPNSTGNTSYSENGTQTEPSPNGNSTEDVSTESNNGNEE 541
QY 861 VOEKLKFTASYGLDVSIFNMDGTIELRPSGEVINKNLS 903
Db 542 IKQENE-----LDPDKVVEPEKTELELR-----NVSD 568
RESULT 9
US-08-923-992A-10
; Sequence 10, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: NO. 6280738-Iga Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.01400001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-3540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

RESULT 10
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: LYNN Doucette-Stamm et al

Db 2637 HRLNEKAP--QOSTIATPNSTIEERQASAKLOEVLAKAKIAKIDKQTNDDVKTYYNG 2695
QY 759 DGTIELPSPSEVKKKNNKADTGAPQNGENKPSKGVSTGVNENK--PADSL 816
Db 2696 IAEIENILPATTVKADKADVAKEQK-----LQINSNDEATTEKLVASDNL 2745
QY 817 PEAREKPVKPNSTNGKMLNPNVNGSDPMLD--PALEAP-ADVPOEKLK 867
Db 2746 NHVETTNQAIADPTNOVNVEKNKGITGIRTIDQPLVVKPTAKSKIESAVEK 2799

RESULT 11
US-08-923-992A-4
; Sequence 4, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438, 0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1104 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-4

Query Match 3.9%; Score 185.5; DB 4; Length 1104;
Best Local Similarity 18.9%; Pred. No. 3.9e-05;
Matches 162; Conservative 131; Mismatches 319; Indels 245; Gaps 39;

QY 166 DEKEHDEVEPTHP---ESDERENHAGLNPASDNLKPKSTDETEEEAEDTDEAIP 222
Db 12 DSVKTEVAAPKPPSMAQDQGNSSSELETRMIPTHIKKAVEPEVKTAGETSAT 71
QY 223 GTPSIRONAMETITGLKSSL---LLGTVKNNITSAEY-----DSLALKESOPAPIQG 273
Db 72 DTGRRKQLOQWNNKANDVNTILSHQENRFTKIDETNDSALLEL-----120
QY 274 PQIGQPTLPNLSLATPSIPNGTSHKEHEGYGFDANRIAEDESQVMSHGDS--331
Db 121 -----ENQFNETNLLRIKHEVEKHNKP-----NQKTLKQSDTKVDSLNIKEL 167
QY 332 NYHFFKND-----LTEEIQKAAQKHEEVKTSNGLDSLSSHEQDYPGNANEMKOLDK 384

Db 168 NHQSQVEAMAEQAGITNEDKDSMLKTIEDIRKQAQADKKEDAE-----VKVREELGK 221
QY 385 KIEKIAIGIMQY--GYRESIVVKNKENAIIYPSGDHHDPI-----426
Db 222 LFSSTKAGLQOOIOEHVAKET---SSEET---QKVDSHYANSLQNTLAQKSLSEELDKATT 275
QY 427 DEHPVPGIGHSHSNYELFKPEEGVAKKGNKVY--TGEEELNVNLLKNSTFNQNFLLA 484
Db 276 NEQATQVKNQFLENAQKLKEIPLIKETNVLKXKAMSSELEQVEKELHNSKANLQDLYA 335
QY 485 NQGRVSEFPPELEKLGINMLVLITPDQKLVKGVGEGVGN1-ANFELDQPYL 543
Db 336 KSKEIVR-----EYEGKL--NQSKNL--PELKQLEEEAHSKLQVVEFRKFKTSEQVT 386
QY 544 PGOTFKYTIASKD-----YPE--VSYDG--FTVPTSLAYKMASTOTIYPPHAG 588
Db 387 PKKVRKDLAANENNOQKIETVSPENITYEGEDVFTVTA---KSDSKT---TLDFS 439
QY 589 DTYLRVNPQFAVPRGTDAVRFVDFEFGNAYLENNYKVGKLPKIPKLNQGTTRT-----643
Db 440 DLLTKYNPSVS-----DRISTNYKTNT---DNHKAETITKLNKLNQSTVTLKAD 488
QY 644 -AGNKIPVTFMANAYLQNSTYIIVEVPILEKENOTDKPSILPQPKRNKAQNSKLDKRV- 701
Db 489 DSGNVVETK-----TITVQKKEK-----QVPKTPQKHKSKEQNV 526
QY 702 EEPKTSKVEKELSETGN-----STSNSTLEEVPTV-----DPVQE 738
Db 527 QEPKNDKQLOELIKSAQOELEKLEKAIKELMEQPEIPSPNPEYGIQKSIWESKEPIQE 586
QY 739 KVAKF-----AESYGMLENVLFNMDGTIELYPSGVEIKNNMADTFGEAPOG 786
Db 587 AITSFNKIIGDSSSKSYTEHYFNKSHFNMYOLHAQN-----EILTKVQVYNNKYPDN 641
QY 787 NGENKPSNGKVS-----GIVENQPT-----NK-----PADSLPEAPNEKP 824
Db 642 AEIKKIFESDMKRTKEDNYGSLNDALKGYFERYFLTPFNKIKQIVDDLDKVKVQDQAP 701
QY 825 VKPENTDNGMLNPEGNVGSDDPMDLPALEA-----PAVDPOEKLKFTASYGLG 875
Db 702 I-PENSE-----MDQAKEKAKIAVSKYMSKVLGVDGVHQLKKNHSKIVD 744
QY 876 L-----DSVIFNMD 884
Db 745 LKLEAIKQOTIFDID 761

RESULT 12
US-08-790-912-3
; Sequence 3, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997

CLASSIFICATION: 424
PRIOR APPLICATION DATA: US 60/026,939
FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-4U1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1964 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-3

Query Match 3.98; Score 184.5; DB 2; Length 1964;
Best Local Similarity 21.38; Pred. No. 0.00011;
Matches 165; Conservative 103; Mismatches 269; Indels 237; Gaps 41;

QY 156 KNAQDDSKPDDEKE--HDEVSEPTHPESDEKENHAGLNPSADNLYKPTSTDEETEERAE 213
DB 51 KYADSELSESKQLVYD---IPTVENDDETYL-----VYKLSNQ----- 92

QY 214 DTTDAEPTGPSIRQNAETITGLKSLLLGTNDNTISAEVDSLLALLKESQAPATQ 273
DB 93 ----LAELPNTGS--KNERQALVAGASLAALGILIFAVSKKKVKNKTVLHLVLVAGMGNG 146

QY 274 PQIGOPTLPNNSLATPSPSLPINPOTSHK-----HEEDGYGFDAENRIIAEDSGFVMSHG 329
DB 147 VLVSVALENHLLNNTDYEL---TSGEKLPKPEISGYTYIGYKGGKTSDEFSN- 202

QY 330 DSNHYFFKKDLTEEQIKAAQKHEEVKTSNGLDLSLSEHDYPCNAKEMKDLKKIEK 389
DB 203 -----QEKSAATPKQKVDYN---VTFNFDHPSTVQAIQEQTPVSTK 244

QY 390 TAGIMQYGVKRESIVVKNENAIYPSGDHHADPIDHKPVGIGHSHSNYELFKPEG 449
DB 245 PTEV-----QVVEKPFSTELINPRKEKQSSDSQOLA-----EKKNET-KREEK 289

QY 450 VAKKEGKVTY---GELTNVNV-----LLNSTN-----NFTLANGKRVSF 492
DB 290 ISPREKTVNTLNPODEVLSQLNKPELLYREETITETKIDFQEEIQENPDLAGGTVRV-- 347

QY 493 SPPPELEKKGAGNM-LVKLIIT-----PDGKVLKVSQK--VFGE----- 528
DB 348 ---KQEGKLGKVEIVRIFSVNKEEVSREIVSTTAPSPRIVEKGYKTVQVKEQPET 403

QY 529 GV-----GNANFELDQPYLPQGTFFYKASKDYPEVSDGTFTVPTSLAYKMASOTI 581
DB 404 GVEHRDVQSGAIVEPAI-QPELP-----EAVVSDKGEVQV-----TLPEAVV----- 446

QY 582 FYPPHAGDTYLRVNPQ-----FAVPGKTDALVRVDFEFGNAYLENNYKVGELK--LPK 635
DB 447 ---TKGET--EVQPESTVTVSDKGEPEQVAPLPEYKGN-----IEQVKPETPVK 493

QY 636 L-NQGTETAGNKIPVTMANAYLDNQSYIIVEVPILEKENQTDKPSILPQKRN---KA 691
DB 494 TREOQPEKT--EEVPV-----KPTETVPVNEGTEGTSI--QEAENPVQPA 537

QY 692 QNSKLDKVEPKTSKY-----EKKLSETGNSTNSLTLEVPVTD 734
DB 538 EESTNSEKVSPTDTSSENTGEVSSNPSTTSVGSNKPENHNSKNENSEKTVVEVP-VN 596

QY 735 PVQEKVAKFAESYGMKLENVLNNDGTIELPLSPSEVIKKNWADTGEAPQNGENKSE 794
DB 597 P-----NEGTV-----GTSNOETEKFPVQPAETQTN 623

QY 795 NGKV---STGTVENQPTENKP--ADSLPEAPNEKPKVPKPS-----TDNGMLNPE 839
DB 624 SGKIANENTGEVSNKPSDSKPPVEESNQPKNGTATKPSNGNTTSNGQTEPE 677

RESULT 13
US-08-790-912-2
Sequence 2, Application US/08790912
Patent No. 5976542
GENERAL INFORMATION:
APPLICANT: Weiser, Jeffrey N.
APPLICANT: Plaut, Andrew G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: 1601 Market Street, 36th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08790,912
APPLICATION NUMBER: US/08790,912
FILING DATE: 29-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,939
FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-4U1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-2

Query Match 3.98; Score 184.5; DB 2; Length 2052;
Best Local Similarity 21.38; Pred. No. 0.00012;
Matches 165; Conservative 103; Mismatches 269; Indels 237; Gaps 41;

QY 156 KNAQDDSKPDDEKE--HDEVSEPTHPESDEKENHAGLNPSADNLYKPTSTDEETEERAE 213
DB 124 KYADSELSESKQLVYD---IPTVENDDETYL-----VYKLSNQ----- 165

QY 214 DTTDAEPTGPSIRQNAETITGLKSLLLGTNDNTISAEVDSLLALLKESQAPATQ 273
DB 166 ----LAELPNTGS--KNERQALVAGASLAALGILIFAVSKKKVKNKTVLHLVLVAGMGNG 219

QY 274 PQIGOPTLPNNSLATPSPSLPINPOTSHK-----HEEDGYGFDAENRIIAEDSGFVMSHG 329
DB 220 VLVSVALENHLLNNTDYEL---TSGEKLPKPEISGYTYIGYKGGKTSDEFSN- 275

QY 330 DSNHYFFKKDLTEEQIKAAQKHEEVKTSNGLDLSLSEHDYPCNAKEMKDLKKIEK 389
DB 276 -----QEKSAATPKQKVDYN---VTFNFDHPSTVQAIQEQTPVSTK 317

QY 390 TAGIMQYGVKRESIVVKNENAIYPSGDHHADPIDHKPVGIGHSHSNYELFKPEG 449
DB 318 PTEV-----QVVEKPFSTELINPRKEKQSSDSQOLA-----EKKNET-KREEK 362

679 KPSILPQFKRNKAQENSKLDEKV-EETPKTSEKVEKEKLSEIGN-----STSNSTLEEVF 731